

THÈSE

présentée

DEVANT L'UNIVERSITÉ DE RENNES I

pour obtenir

le grade de : **DOCTEUR DE L'UNIVERSITÉ DE RENNES I**

Mention : Traitement du Signal et Télécommunications

par

Vincent Auvray

Équipe d'accueil : VISTA (IRISA, RENNES)

École doctorale : MATISSE

Composante universitaire :

Titre de la thèse :

Estimation et segmentation des mouvements transparents dans des séquences d'images fluoroscopiques avec application au débruitage

COMPOSITION DU JURY :

M.	Jean-Pierre	Le Cadre	Président
M.	Til	Aach	Rapporteurs
M.	Denis	Pellerin	
M.	Nikos	Paragios	
M.	Patrick	Bouthemy	
M.	Jean	Liénard	

Remerciements

Ces travaux ont été menés dans un milieu hétérogène : ils ont été effectués pour moitié à l'Institut de Recherche en Informatique et Systèmes Aléatoires, au sein de l'équipe Vista, et pour moitié chez General Electric Healthcare, au sein de l'équipe Advanced Medical Applications.

Je remercie avant tout Til Aach et Denis Pellerin d'avoir accepté, après avoir encadré des thèses sur l'estimation des mouvements transparents, d'être les rapporteurs attentifs de la mienne. Je tiens également à remercier Nikos Paragios et Jean-Pierre Le Cadre d'avoir examiné ce travail et complété le jury.

Je remercie mon maître de thèse industriel, Jean Liénard, pour sa gentillesse et sa disponibilité. Et je souhaite remercier tout particulièrement Patrick Bouthemy pour la qualité de son encadrement, la pertinence de ses conseils et l'investissement formidable dont il a fait preuve tout au long de ces trois années.

Avant de glisser vers les remerciements plus privés, je souhaite remercier les "amis de travail", au premier rang desquels Vincent Bismuth, mon alter-égo à l'écoute attentive, aux suggestions pertinentes et aux anecdotes illuminant des matinées de travail. Ce qui me permet de remercier toute l'équipe des pauses-café : Adiila, Bene, Muriel, JB, Steph, Richard, Benjamin, Guillaume, Florian, Gero et Reynald. Je n'oublie pas de dynamisme et les remises en question salvatrices de Rêgis, et plus largement la bonne ambiance de l'équipe AMA. Enfin, je remercie Mark Grabb pour m'avoir accueilli dans son équipe au sein du centre de recherche de GE quelques semaines, ainsi que François Nicolas et François Kotian pour avoir permis qu'elle soit un modèle de Cifre, équilibrée entre les mondes industriels et académiques.

Je remercie également toute l'équipe rennaise pour m'avoir immédiatement intégré malgré mon rythme pendulaire. Que mes colocataires successifs et ceux qui n'auront pas eu ce bonheur en soit remerciés : Thomas B., Anne, Élise, Jérôme, Nico N., Thomas V., Alex, Gwen, Aurélie, Vincent, Nico G., Patrick H et Ivan. Merci à Huguette pour son efficacité et sa gentillesse, merci aux permanents pour les échanges décontractés, et particulièrement à Charles pour son intérêt pour le débruitage.

Et je conclus par une salve de remerciements "que" personnels. Merci à vous, parents patients ; merci à toi, frangin artiste ; merci à vous, grands-parents ouverts ; merci à toi, famille super ; grazie a voi, famiglia ospitale. Merci à toi, zeliste ; merci à vous, couple sniper ; merci au basque et sa québécoise ; merci à toi, barbu sportif ; merci à toi, canadien alpiniste ; merci à toi, toulousain déconnecté ; merci à vous, les passionnés du pneu et merci à vous, les polonais sportifs. Merci à toi Aachen ; merci à toi le fan de gouda et kebab ; merci à toi mister Doudard ; merci à toi, blonde positive ; merci à toute la smala ; merci à vous les Erasmus italiennes. Et surtout merci à toi, mon Elena.

Et bien évidemment, merci à toi, lecteur, de t'attaquer à ce lourd document.

Table des matières

Introduction générale	11
I Contexte applicatif et enjeux	15
1 L'imagerie X interventionnelle	17
1.1 Diagnostic et traitement de maladies cardiovasculaires	17
1.1.1 Maladies vasculaires	17
1.1.2 Diagnostic et traitement	19
1.1.3 Système d'imagerie cardiovasculaire par rayons X	20
1.2 Formation des images	21
1.2.1 Physique des rayons X	21
1.2.1.1 Principe	21
1.2.1.2 Propriétés du rayonnement	22
1.2.2 Mesure des images et prétraitements	24
1.2.3 Caractéristiques des images et objectifs de la thèse	26
1.3 Des images transparentes	29
1.3.1 Modèle des couches	29
1.3.2 Simulation d'images réalistes	30
1.3.2.1 Génération de cartes d'atténuation de référence	30
1.3.2.2 Synthèse de séquences réalistes	31
1.4 Conclusion	33
II Estimation de mouvements transparents	35
2 Estimation du mouvement avec transparence ; état de l'art	37
2.1 La problématique de la transparence : vers des estimateurs dédiés	38
2.1.1 Adaptation partielle d'estimateurs de mouvements classiques	38
2.1.2 Une première modélisation complète de la transparence par généralisation du flot optique	41
2.1.2.1 Principe	41
2.1.2.2 Méthode de résolution	43
2.2 Estimation de mouvements transparents dans l'espace de Fourier par généralisation de la conservation de l'intensité	44
2.2.1 Principe	44
2.2.2 Méthodes de résolution	45

2.3	Estimation de mouvements transparents dans l'espace direct	47
2.3.1	Principe	47
2.3.2	Méthodes de résolution	51
2.4	La problématique de la séparation des couches	55
2.4.1	Formulation dans l'espace des fréquences	56
2.4.2	Filtrage spatial dans l'espace direct recalé	56
2.4.3	Résolution directe du système	60
2.4.4	Séparation des couches appliquée aux images fluoroscopiques	61
2.5	Conclusion	61
3	Estimation du mouvement en situation de bi-transparence pour les examens par rayons X	63
3.1	Équation de contrainte du mouvement transparent avec modèles paramétriques	63
3.2	Estimation multirésolution du mouvement transparent	65
3.2.1	Linéarisation et résolution	66
3.2.1.1	Au premier ordre	66
3.2.1.2	Au second ordre	68
3.2.2	Formalisme multirésolution	70
3.3	Estimation robuste du mouvement transparent	71
3.3.1	Intérêt d'une estimation robuste	71
3.3.2	La fonction robuste choisie	71
3.3.3	Minimisation robuste par la méthode des moindres carrés pondérés itérés	73
3.4	Initialisation	74
3.4.1	Initialisation par la méthode du simplex	75
3.4.2	Mise en correspondance par blocs pour la bi-transparence	75
3.4.3	Séparation des champs de déplacement	76
3.4.3.1	Clustering et estimation robuste.	76
3.4.3.2	Transformée de Hough.	78
3.5	Résultats	82
3.5.1	Résultats synthétiques	82
3.5.2	Traitement de séquences d'images réelles	92
3.5.2.1	Cas de condition normale de bi-transparence	92
3.5.2.2	Cas aux limites de validité des hypothèses	96
3.6	Conclusion	99
4	Estimation du mouvement en situation de bi-transparence distribuée	103
4.1	La bi-transparence distribuée	104
4.1.1	Définition	104
4.1.2	Une segmentation en couches	105
4.2	Estimation unifiée des mouvements des couches	106
4.2.1	Résolution pratique du problème	106
4.2.2	Initialisation	107
4.2.2.1	Estimation des modèles de mouvement sur chaque région	108

4.2.2.2	Attribution des modèles affines estimés sur les régions aux couches	109
4.2.2.3	Extraction des modèles à étiquetage connu	109
4.2.2.4	Initialisation de l'appariement	111
4.3	Résultats	111
4.3.1	Résultats sur séquences synthétiques	111
4.3.1.1	Synthèse d'images réalistes en situation de bi-transparence distribuée	111
4.3.1.2	Expériences sur séquences synthétiques	113
4.3.2	Traitement de séquences réelles	116
4.3.2.1	Dispositif pratique	116
4.3.2.2	Résultats	117
4.4	Conclusion	123
5	Estimation jointe du mouvement et de la segmentation en couches en situation de bi-transparence distribuée	125
5.1	Un problème joint d'estimation et de segmentation par le mouvement . . .	125
5.1.1	Approche	125
5.1.2	Formalisation markovienne	126
5.1.3	Minimisation de l'énergie globale	128
5.2	Initialisation	130
5.3	Gestion du nombre de couches	132
5.3.1	Mécanisme de réduction du nombre de couches	132
5.3.2	Utilisation pratique de l'ajout de couche	132
5.4	Comportement sur les blocs mono- et multi-couches	134
5.4.1	Comportement sur un bloc mono-couche	134
5.4.2	Comportement sur un bloc multi-couche	136
5.4.3	Pertinence de l'adaptation du schéma d'estimation-segmentation pour détecter les configurations hors bi-couches	137
5.5	Résultats expérimentaux	138
5.5.1	Traitement de séquences synthétiques	138
5.5.1.1	Choix des métriques	138
5.5.1.2	Simulations	139
5.5.2	Résultats sur séquences réelles	143
5.6	Conclusion	147
	Synthèse de la partie II	149
	III Débruitage compensé en mouvement avec transparence	155
	Introduction de la partie III	157
6	État de l'art du débruitage de séquence d'images	159
6.1	Filtres pondérateurs	159

6.2	Filtres séquentiels.	161
6.3	Filtres d'ordre	162
6.4	Restauration bayésienne	163
6.5	Filtrage dans des espaces transformés	163
6.6	Filtrage par EDP	164
6.7	Le problème de la mesure de la qualité du débruitage	165
6.8	La question du mouvement	166
7	Débruitage temporel compensé en mouvement de séquences d'images transparentes	167
7.1	Filtre temporel récursif non compensé en mouvement	167
7.1.1	Principe	167
7.1.2	Comportement sur régions stationnaires	168
7.1.3	Adaptation du gain	169
7.1.3.1	Filtrage adaptatif	169
7.1.3.2	Réglage pratique du gain	169
7.1.3.3	Application pratique.	171
7.2	Compensation des mouvements transparents	173
7.2.1	Principe	173
7.2.2	Limitations de la compensation des mouvements transparents	174
7.2.3	Résultats	175
7.3	Conclusion.	176
8	Débruitage temporel hybride de séquences d'images transparentes	179
8.1	Filtre temporel hybride	179
8.1.1	Principe	179
8.1.2	Quatre configurations locales différentes	180
8.1.3	Une cinquième configuration	180
8.2	Filtrage linéaire	181
8.3	Détection de la configuration	182
8.3.1	Critère de décision.	182
8.3.2	Fonction de décision	185
8.4	Résultats à mouvements connus	188
8.4.1	Traitement de la séquence test	188
8.4.2	Étude des cartes de poids	190
8.5	Résultats avec estimation des mouvements transparents	194
8.6	Résultats sur séquences réelles	196
8.7	Conclusion	200
9	Débruitage spatio-temporel compensé selon les mouvements transparents	203
9.1	Filtres spatio-temporels compensés en mouvement	203
9.1.1	Principe	203
9.1.2	Résultats sur séquence synthétique	204
9.2	Filtres spatio-temporels hybrides	206
9.2.1	Principe	206

9.2.2	Résultats sur séquences synthétiques	208
9.3	Conclusion	211
Conclusion générale et perspectives		213
A	Séparation des couches transparentes sur images médicales	221
A.1	Médian temporel sur séquences stabilisées	221
A.1.1	Tests synthétiques	221
A.1.2	Influence de la qualité de la stabilisation	222
A.1.3	Séparation de la couche statique	223
A.2	Méthode d'échange d'informations au niveau des couches	225
A.2.1	Approche globale	225
A.2.2	Approche locale	226
A.3	Conclusion	227
B	Minimisation du critère de bitransparence linéarisé	229
B.1	Estimation de mouvements paramétriques par la technique de Gauss-Newton	229
B.1.1	L'équation de contrainte du mouvement transparent impliquant des modèles paramétriques	229
B.1.2	Un formalisme unique pour les deux configurations de transparence .	230
B.1.3	Minimisation par la technique de Gauss-Newton	231
B.2	Linéarisation au premier ordre	231
B.2.1	Algorithme de minimisation associé à l'équation de contrainte du mouvement avec transparence linéarisée au premier ordre	232
B.3	Minimisation du critère linéarisée au second ordre	234
B.3.1	Développement limité du critère au second ordre	234
B.3.2	Calcul de la direction de descente	235
B.3.3	Homogénéisation de la direction de descente	236
B.3.4	Choix du pas	237
B.3.5	Algorithme de minimisation associé à l'équation contrainte du mou- vement avec la transparence linéarisée au second ordre	237
C	Ajout de couches	239
C.1	Détection de blocs mal étiquetés.	239
C.2	Choix de la couche erronée.	241
C.3	Estimation du modèle affine de la nouvelle couche.	243
C.4	Utilisation pratique de l'ajout de couches	244
D	Détection des configurations mono- et multi-couches	247
D.1	Traitement des blocs en configuration mono-couche	247
D.1.1	Détection des blocs suspects	247
D.1.2	Intégration de la détection mono-couche	250
D.2	Détection des blocs en configuration de multi-couches	251
D.2.1	Détection des blocs suspects	251
D.2.2	Intégration de la détection de configurations de multi-couches	252
D.3	Discussion sur la pertinence de la détection de configurations	252

E	Traitement d'images vidéos transparentes	257
E.1	Une séquence de bi-transparence	257
E.2	Une séquence synthétique de bi-transparence distribuée	259
E.3	Une séquence de laboratoire de bi-transparence distribuée	261
E.4	Une séquence réelle de bi-transparence distribuée	261
F	Étude de la limite de débruitage du filtre temporel récursif avec compensation des mouvements	265
F.1	En situation monocouche	265
F.2	En situation de bi-transparence	267
F.3	En situation de transparence à n couches	269
G	Calcul des coefficients du filtrage hybride	273
G.1	Dans l'hypothèse (H_0)	273
G.2	Cas de l'hypothèse (H_1)	274
G.3	Cas de l'hypothèse (H_2)	275
G.4	Cas de l'hypothèse (H_3)	275
	Liste des algorithmes	276

Introduction générale

L'imagerie médicale par rayons X ouvre de nombreuses possibilités de diagnostic et de thérapie. Le médecin peut notamment éviter des examens chirurgicaux lourds en choisissant d'exploiter des images acquises par cette modalité, qui lui permettent de visualiser l'anatomie du patient.

C'est ainsi que son utilisation est très largement répandue dans le domaine cardiovasculaire : à partir de *vidéos* acquises par rayons X, le clinicien peut examiner le réseau vasculaire du patient et le cas échéant décider de mener une intervention. Une des options qui lui est offerte est d'effectuer cette intervention de manière minimalement invasive, en guidant son geste par les vidéos de l'anatomie acquises par rayons X en temps réel.

Cependant, les rayons X sont nocifs pour la santé, que ce soit celle du patient ou celle du personnel médical, de sorte que l'exposition doit être aussi limitée que possible. Il en résulte des séquences d'images de nature difficile : elles sont très bruitées et peu contrastées. Dans ces conditions, le traitement de l'image apporte une aide précieuse au médecin en mettant en évidence l'information pertinente contenue dans les images, tout en rendant leur observation intensive aussi agréable que possible.

L'imagerie médicale par rayons X mène donc à des problématiques de traitement de l'image importantes en terme d'applications. Elles constituent également des défis scientifiques passionnants du fait de la difficulté des images considérées. Ils sont d'autant plus difficiles à résoudre qu'il est nécessaire pour de nombreuses applications de travailler en temps réel, à des taux d'acquisition élevés (30 images par seconde en général).

Nous nous intéressons dans cette thèse à **l'estimation des mouvements anatomiques** présents dans ces images. Cette tâche est complexe à plus d'un titre :

- **Les images sont transparentes.** La physique de formation des images, basée sur l'*atténuation* des rayons X lors de leur passage dans l'anatomie, implique une transparence multiplicative de l'image, qui est transformée en transparence additive (plus facile à manipuler) par un opérateur log. Lorsque deux objets se chevauchent, on n'assiste pas à l'occultation de l'un par l'autre mais à la superposition de leurs niveaux de gris. Par conséquent, les méthodes classiques d'estimation de mouvement ne sont pas applicables puisqu'elles postulent la conservation de l'intensité d'un point le long de sa trajectoire dans l'image.
- **Les images sont bruitées et peu contrastées.** Du fait du rayonnement limité utilisé, les images sont entachées d'un fort bruit quantique. Comme d'autre part les

objets d'intérêt sont peu contrastés (leur RCB^1 s'étage entre 0,5 et 3), les données sont difficiles à exploiter.

- L'application majeure visée, le débruitage compensé en mouvement, implique à terme une **estimation des mouvements en temps réel**. Si nous ne nous fixons pas cet objectif dans cette thèse, nous nous interdisons cependant d'utiliser des approches trop coûteuses en temps de calcul. Nous souhaitons qu'une implantation en temps réel des méthodes que nous proposons reste raisonnablement envisageable.
- **Les mouvements considérés sont mal définis** : ce sont les projections $2D$ de mouvements $3D$, qui sont pour certains complexes. Si le mouvement respiratoire est très régulier, les battements cardiaques sont autrement plus complexes par exemple. Pour contraindre suffisamment le problème, nous les modéliserons par des modèles de mouvement affines. Cette approximation est acceptable puisqu'elle sera utilisée avec une méthode de segmentation automatique de l'image qui permettra de sursegmenter les anatomies au mouvement trop complexe pour être représenté par un modèle affine unique.

Nous avons pris le parti de nous concentrer sur l'estimation des mouvements anatomiques, en excluant notamment le suivi des outils interventionnels. Ils ont en effet la particularité d'occuper une faible surface de l'image, de sorte que nous pensons qu'il n'est pas réaliste de calculer leur mouvement dans un formalisme global. Deux stratégies sont envisageables pour estimer leurs mouvements :

- ou bien calculer les mouvements des organes avec lesquels ils se déplacent. Les guides, cathéters et stents sont en effet solidaires des artères dans lesquels ils se situent, de sorte que les mouvements des artères permettent d'accéder aux mouvements des outils (quand le médecin ne leur insuffle pas de déplacement propre) ;
- ou bien de développer des approches dédiées, intégrant autant d'*a priori* possible sur la géométrie de l'outil, le type de mouvement attendu, etc [Baert 03].

En fait, nous constaterons que les méthodes que nous proposons permettent dans certaines situations d'estimer aussi le mouvement des outils interventionnels.

Les applications de l'estimation des mouvements anatomiques sont multiples.

- L'angiographie soustraite est une modalité qui permet de mettre en évidence le réseau vasculaire en soustrayant un masque injecté à l'image acquise en temps réel. Pour qu'aucun artefact ne corrompe l'image finale, il est nécessaire de compenser les mouvements ayant eu lieu entre l'acquisition du masque et l'acquisition courante [Meijering 99], ce qui nécessite le calcul des mouvements anatomiques [Mittal 04].
- La compression des images acquises pour archivage [Nzomigni 93] est plus efficace quand une compensation des mouvements a eu lieu, car la redondance temporelle est accrue.
- Nous pouvons également envisager l'extraction de métriques liées au mouvement, notamment dans le cycle cardiaque, ce qui permettrait de quantifier d'éventuelles insuffisances. Enfin, l'estimation des mouvements anatomiques peut aider au suivi d'objets d'intérêt, dans le but de les réhausser par exemple [Baert 03].

L'application essentielle que nous visons est la mise au point d'un **filtre réducteur de bruit** filtrant temporellement "dans le sens du mouvement". Nous entendons par là qu'il

¹Rapport du Contraste de l'objet sur l'écart-type du Bruit.

doit être capable de compenser les mouvements anatomiques, ce qui permettra de réduire fortement le bruit sans introduire d'artefacts (comme la réduction du contraste des objets en mouvement).

Plan du document

Ce document est structuré en trois parties selon le plan suivant :

Partie 1 - Contexte applicatif et enjeux

Nous présentons dans une première partie le **cadre clinique**. Nous exposons en particulier les **propriétés physiques particulières des images** acquises par rayons X. Elles sont entachées d'un bruit important, d'origine quantique essentiellement, et sont corrompues par un "voile" dû au rayonnement indirect (le diffusé).

Leur propriété fondamentale pour ce qui nous concerne est la *transparence*. Ceci impose le développement de méthodes d'estimation du mouvement prenant explicitement en compte la physique de formation des images.

Cette partie se conclut par une méthode de simulation d'images cliniques réalistes, à vérité-terrain connue (mouvements, bruit, diffusé). Cela nous donnera les moyens d'évaluer quantitativement les performances des estimateurs que nous allons développer.

Partie 2 - Estimation des mouvements transparents

Nous présentons tout d'abord une **revue des différentes méthodes d'estimation de mouvements transparents** proposés jusqu'à présent. Nous en concluons que seules des approches s'appuyant sur l'équation de contrainte du mouvement avec transparence (qui est un équivalent à la conservation de l'intensité dans le cas non transparent) sont envisageables dans notre cas.

Pour résoudre le problème, nous ferons deux hypothèses majeures :

- *en chaque point de l'image, au plus deux couches sont présentes*. Si la réalité est bien sûr plus complexe du fait de la nature tridimensionnelle du problème, cette hypothèse est visuellement confortée par l'observation des examens réels. En outre, nous montrerons que les méthodes que nous proposerons sont capables de retenir les deux couches dominantes dans des situations "tangentes".
- *les mouvements sont représentés par des modèles paramétriques simples*. Cette hypothèse, qui peut sembler forte au premier abord, est acceptable car nous nous focalisons sur le mouvement des organes, qui en général est régulier. D'autre part, nous étudierons une segmentation automatique de l'image en couches, qui est à même de sursegmenter une zone anatomique en différentes régions si le mouvement est trop complexe pour être expliqué par un modèle paramétrique unique.

C'est dans ce cadre que nous décrivons dans un premier temps une **méthode d'estimation des mouvements en situation de bi-transparence simple** dans le chapitre 3, c'est-à-dire une méthode dédiée aux images constituées globalement de deux couches exactement. Notre approche s'articule autour de trois étapes principales : calcul des vecteurs de déplacements (doubles) par une méthode adaptée de mise en correspondance par blocs,

extraction des deux modèles de mouvement par transformée de Hough, et enfin estimation paramétrique globale des mouvements transparents.

Nous nous intéressons ensuite au cas de séquences d'images en situation de *bi-transparence distribuée*, c'est-à-dire contenant un nombre quelconque de couches globalement, mais au plus deux en chaque point. Le chapitre 4 présente une **méthode d'estimation des mouvements en situation de bi-transparence distribuée à segmentation donnée**, qui garde la même structure que l'approche du chapitre 3.

Le chapitre 5 est lui consacré à l'**estimation des mouvements en situation de bi-transparence distribuée** sans segmentation connue. Nous proposons une estimation et une segmentation jointes du mouvement, par minimisation itérative d'une énergie globale de type markovienne.

Chacune de ces trois méthodes est appliquée aux images test réalistes, ainsi qu'à un grand nombre d'exams réels, avec des résultats satisfaisants.

Partie 3 - Débruitage compensé en mouvement avec transparence

La troisième partie est consacrée à l'utilisation de cette information de mouvement à des fins de débruitage. Le chapitre 6 présente un rapide **état de l'art des filtres de réduction de bruit appliqués à des séquences d'images**. Nous choisissons d'explorer deux pistes : le filtrage purement temporel, et le filtrage spatio-temporel.

Du fait de la transparence, il n'est pas possible de procéder à la compensation de mouvement de manière habituelle en filtrant le long de la trajectoire de chaque point. Nous proposons dans le chapitre 7 une méthode de compensation de mouvement transparent, que nous appliquons à un **filtre purement temporel**. Cette approche se heurte à une limite asymptotique de débruitage réduite. C'est pourquoi nous proposons une alternative permettant des réductions de bruit plus significatives dans le chapitre 8, tout en conservant pour une large part les qualités de conservation des contrastes de l'approche compensée en mouvements transparents. Le filtre en question, nommé *filtre hybride temporel*, décide en chaque pixel si la compensation de mouvements transparents y est pertinente ou non. Nous appliquons le même type d'approche dans le chapitre 9 à un filtre **spatio-temporel**. Il permet un débruitage beaucoup plus important que des méthodes temporelles pures, mais modifie l'esthétique de l'image, ce qui gêne les médecins. Même si pour l'heure ce type de filtre entraîne des artefacts mal maîtrisés, nous montrons qu'une compensation des mouvements transparents de type "hybride" améliore ses performances.

Enfin, la conclusion générale présente une synthèse de cette étude et quelques perspectives liées à ces travaux.

Première partie

Contexte applicatif et enjeux

Chapitre 1

L'imagerie X interventionnelle - Cadre clinique et propriétés des images

Cette thèse porte sur le traitement des images issues d'examens cardiovasculaires par rayons X. Les procédures cliniques de ce type sont effectuées très fréquemment afin de diagnostiquer et de traiter les maladies cardiovasculaires.

Nous présentons dans ce premier chapitre le cadre clinique, ainsi que les propriétés très particulières des images issues de ces examens, ce qui permet d'introduire la problématique de la thèse. Nous insistons notamment sur le fait que les images que nous considérons sont transparentes.

Enfin, nous détaillons une méthode de génération d'images test permettant de mesurer quantitativement la précision de nos méthodes d'estimation de mouvements et de débruitage dans une situation réaliste.

1.1 Diagnostic et traitement de maladies cardiovasculaires

D'après l'INSEE, 30% des décès en France sont causés par des cardiopathies. C'est la première cause de mortalité en Europe. Cela dit, grâce au développement de méthodes de diagnostic et de traitement efficaces, le taux de mortalité correspondant à ces maladies a diminué en France de 30% entre 1989 et 2000¹.

Cette première section décrit les maladies vasculaires et leur impact, avant de présenter les examens permettant de diagnostiquer les pathologies et de les traiter. Nous nous attardons sur la cardiographie, l'angiographie et l'angioplastie, qui nous intéressent dans cette thèse.

1.1.1 Maladies vasculaires

Le réseau vasculaire fournit l'organisme en oxygène (Fig.1.1). Il circule du coeur aux organes en empruntant les artères, puis revient par les veines.

Le coeur est composé de quatre cavités (deux *ventricules* et deux *oreillettes*, Fig.1.2) qui

¹Toujours d'après l'INSEE.

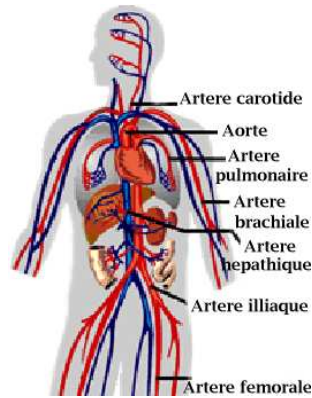


FIG. 1.1: Réseau vasculaire

permettent une double circulation (vers les organes d'une part et les poumons d'autre part). Son muscle, le *myocarde*, est lui-même irrigué par l'intermédiaire des *artères coronaires*. Leur rôle est crucial car l'alimentation en oxygène du myocarde doit être parfaite pour permettre au cœur d'assurer ses quelque 80000 battements quotidiens. Si l'organisation des artères coronaires varie grandement selon les individus, on peut toujours identifier les deux principales : la coronaire droite, et la gauche, qui se divise ensuite en la coronaire interventriculaire antérieure et la coronaire circonflexe.

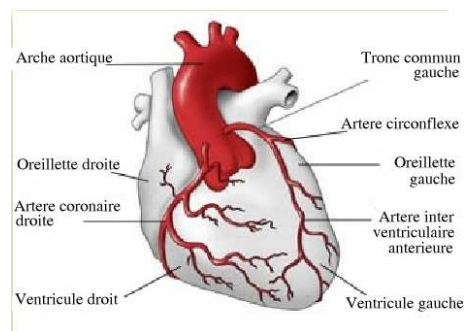


FIG. 1.2: Le cœur, ses deux oreillettes et ventricules et les artères coronaires principales.

Nous pouvons distinguer deux grandes causes de maladies cardiovasculaires.

- **Les sténoses.** Il arrive que les artères retrécissent ou s'obstruent par dépôt de lipides. Une *plaque athéroscléreuse* se forme alors et rend la circulation plus difficile. On parle de formation de *sténoses*.

Elles sont particulièrement dangereuses quand elles touchent les artères coronaires : il s'agit alors de *coronaropathies*. Lorsque l'obstruction est encore légère, le patient ressent une pression thoracique, en particulier pendant l'effort. Il peut également souffrir de brûlures d'estomac, de nausées ou de difficultés respiratoires.

Lorsque la sténose est très développée, les symptômes sont beaucoup plus graves et

vont de l'angine de poitrine (douleur thoracique) à la crise cardiaque (infarctus du myocarde), en passant par la naissance de troubles du rythme cardiaque (arythmies). D'autres artères peuvent également être touchées par ce dépôt de plaque ; on parle alors de *maladies vasculaires périphériques*. Les organes concernés sont essentiellement les jambes, les bras ou les reins. Les symptômes dépendent de la région touchée.

- **Les troubles du rythme cardiaque.** Le rythme cardiaque est contrôlé par des signaux électriques (ce sont eux qui sont mesurés lorsque l'on pratique des *électrocardiogrammes* (ECG)). Des signaux électriques anormaux et dangereusement rapides peuvent mener à une *mort subite cardiaque*. Elle est particulièrement délicate à diagnostiquer car elle ne donne pas lieu à des symptômes clairs. Cette pathologie provient la plupart du temps d'une crise cardiaque antérieure qui a causé la mort ou la mauvaise cicatrisation de tissus.

Les facteurs de risques favorisant l'apparition de maladies vasculaires sont avant tout le tabagisme et une haute pression artérielle. Le cholestérol, l'obésité, le stress, l'absence d'activité physique et les antécédents familiaux ont aussi un impact important.

Ces maladies sont extrêmement fréquentes, et pas seulement dans les pays développés car d'après l'OMS elles deviendront la première cause de mortalité mondiale en 2008. Leur coût social est aussi extrêmement lourd puisque l'American Heart Association évalue le coût pour les seuls Etats-Unis des attaques cardiaques à 54 milliards de dollars en 2004 (33 milliards de manière directe pour les thérapies et 21 de manière indirecte en perte de productivité).

1.1.2 Diagnostic et traitement

En cas de suspicion de pathologie vasculaire, des ECGs au repos, puis à l'effort, peuvent être effectués, en particulier pour recueillir des indices indiquant une obstruction de vaisseaux. Les arythmies cardiaques sont plus difficiles à mettre à jour et peuvent par exemple nécessiter le port d'un *système de Holter*, qui mesure le rythme cardiaque sur 24 heures. L'*échocardiographie* permet de mesurer des indicateurs plus fins tels que la fraction d'éjection (rapport du flux à l'entrée et à la sortie d'une artère). L'*imagerie nucléaire* (PET, SPECT) permet quant à elle, en affichant certains échanges moléculaires marqués par des indicateurs radioactifs, de visualiser la carte d'activité du cœur.

Mais l'examen le plus exhaustif dédié à la visualisation du système vasculaire est l'**angiographie**. Il s'agit d'un examen minimalement invasif qui permet d'afficher sur des images à rayons X l'arbre vasculaire du patient, par injection de produit de contraste. Le praticien introduit un tube spécifique, le *cathéter* à proximité de l'artère qu'il souhaite imager, puis y injecte un produit radio-opaque, le plus souvent de l'iode. Il observe alors grâce à un système de visualisation par rayons X le réseau artériel qui l'intéresse et peut dresser un diagnostic précis sur la présence ou non de sténoses (Fig.1.5).

Les options de traitement sont alors les suivantes :

- **Médicamenteuse.** Même si les médicaments actuels ne peuvent résoudre les problèmes de rétrécissements des vaisseaux, ils peuvent aider l'efficacité coronaire, ce qui peut suffire face à certaines pathologies légères.
- **Pontage.** C'est à l'inverse une procédure très lourde, consistant à greffer un vaisseau sanguin de la jambe ou de la poitrine sur le cœur pour contourner la zone irrémédiablement rétrécie.

- **Cathéterisme.** Les cardiologues peuvent opter pour une solution intermédiaire en terme de lourdeur d'intervention. Il s'agit de déboucher mécaniquement les vaisseaux en manoeuvrant des outils dans le réseau artériel du patient. Il peut s'agir d'*angioplastie* ou d'*athérectomie*. La première consiste en le gonflage d'un petit ballon au niveau de la plaque qui l'écrase et la repousse par pression (Fig.1.3), et est suivie ou non de la pose d'un *stent* (petit ressort) dont la paroi a été traitée chimiquement pour prévenir le redépôt de plaque. Le médecin peut préférer une athérectomie : une lame au bout du cathéter permet alors de découper la plaque et de la récupérer dans un petit réservoir. Quoiqu'il en soit, le praticien visualise son geste par voie des rayons X.

Selon la gravité de la pathologie et l'état de santé du patient, les médecins arbitrent entre ces trois options aux risques et aux pérénalités différentes.

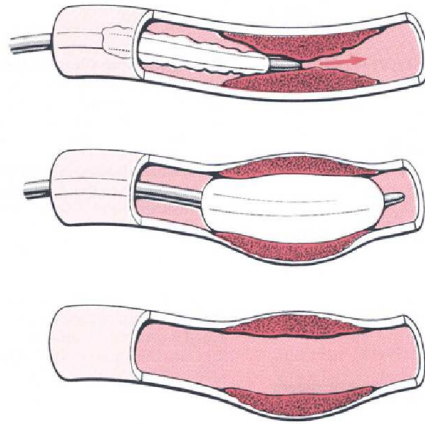


FIG. 1.3: L'angioplastie. Le gonflage du ballon repousse la plaque.

1.1.3 Système d'imagerie cardiovasculaire par rayons X

Dans ce dernier type d'intervention, très répandue, l'imagerie joue un rôle crucial. Le système de visualisation est composé d'une table sur laquelle est allongé le patient, d'écrans portés par un bras télescopique et du bras en C (Fig. 1.4). Ce dernier porte à une extrémité le tube qui émet les rayons X, et à l'autre le récepteur numérique qui mesure les photons X ayant traversé le patient. Le bras en C peut être tourné selon les trois axes pour permettre d'imager l'anatomie dans toutes les angulations nécessaires au diagnostic.

Lors d'un examen, on administre au patient un anesthésique local qui l'engourdit. Il est sous perfusion, et ses signaux vitaux sont mesurés par des électrodes.

Le praticien désinfecte et ponctionne l'artère fémorale au niveau de l'aîne (ou l'artère radiale au niveau du poignet), par laquelle il introduit un fil métallique très fin dans le réseau vasculaire (le *guide*). Il le positionne dans la partie de l'anatomie qui l'intéresse en naviguant dans les artères du patient depuis le point de ponction.

Ce guide sert alors de "rail" sur lequel le médecin s'appuie pour mener ses outils interventionnels à l'endroit à traiter. Il s'agit d'abord du *cathéter*, ce tube qui permet l'injection de

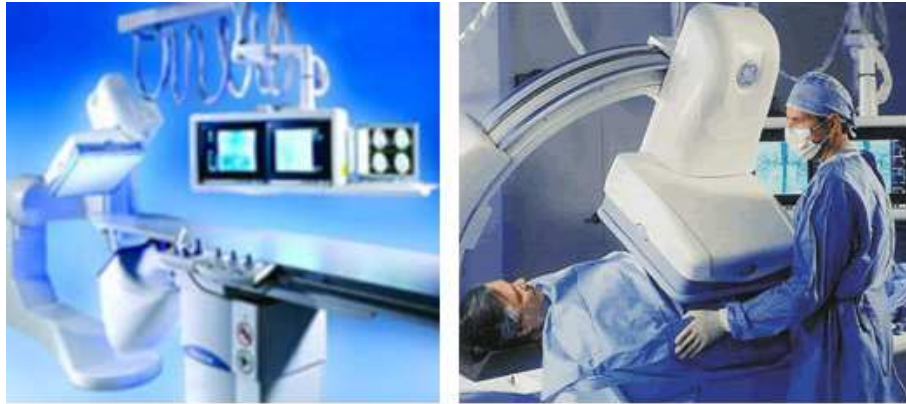


FIG. 1.4: Système d'imagerie par rayons X. (Ici, Innova 4100 de General Electric). A gauche, le système seul, composé de la table, du bras en C portant à une extrémité le tube émetteur de rayons X (en bas) et à l'autre le capteur, et des écrans de visualisation. A droite, le même système lors d'un examen.

produit de contraste radio-opaque rendant les vaisseaux visibles. Les autres outils (stents, ballons, *coils* permettant de boucher des anévrismes) suivent le même chemin.

Le médecin se base pour guider son geste uniquement sur les images affichées sur les écrans qui lui font face, et lui présentent des projections de l'anatomie traversée par les rayons X. Il est donc absolument crucial que ces images soient aussi justes et précises que possible.

Le chapitre suivant présente la physique de formation de ces images, ce qui permettra de comprendre leurs propriétés et de présenter le problème que nous nous sommes posé dans le cadre de cette thèse.

1.2 Formation des images

Les examens par rayons X donnent naissance à des images aux propriétés particulières, ne serait-ce que parce qu'elles s'appuient sur le rayonnement *transmis* à la différence des images vidéo classiques qui mesurent la lumière *réfléchi*e.

Cette partie présente la physique gouvernant ce type d'imagerie et détaille les prétraitements que subit l'image native pour être lisible. Ceci permet d'en venir aux propriétés des images et d'introduire les objectifs de la thèse. Cette présentation ne couvre que les aspects principaux de la formation des images qui importent du point de vue de cette thèse. Des études plus approfondies, qui insistent en particulier sur les mesures *Noise Equivalent Quanta* (NEQ) et *Detective Quantum Efficiency* (DQE) permettant de quantifier à la fois le transfert du signal et du bruit, se trouve dans [Rabbani 87, Spekowius 95, Aach 99].

1.2.1 Physique des rayons X

1.2.1.1 Principe

Les rayons X sont atténués d'une manière différente par les matériaux qu'ils traversent [Macovski 04]. On peut attribuer aux matériaux un coefficient d'atténuation μ_{mat} qui per-

met de relier le flux $\Phi_{out,ideal}$ de photons X sortant d'un objet d'épaisseur d au flux Φ_{in} entrant :

$$\Phi_{out,ideal} = \Phi_{in} \cdot e^{-\mu_{mat} \cdot d} \quad (1.1)$$

Nous faisons l'hypothèse d'un rayonnement monochromatique pour établir ce résultat. Le champ sortant est décrit comme étant idéal car il est en fait entaché de bruit et d'un phénomène de diffusé, comme nous le verrons plus loin.

Ainsi, différents organes réagiront différemment à la même radiation. Le rayonnement ayant traversé les n tissus d'épaisseur d_i et de coefficients μ_i vaudra donc :

$$\Phi_{out,ideal} = \Phi_{in} \cdot \prod_{i \in [1,n]} e^{-\mu_i \cdot d_i} \propto \prod_{i \in [1,n]} e^{-\mu_i \cdot d_i} \quad (1.2)$$

Il est ainsi possible en soumettant le patient à un rayonnement X et, en décomptant localement les photons X l'ayant traversé, de visualiser l'anatomie sous-jacente. Un exemple d'image obtenue est présenté Fig.1.5.

Il n'y a pas d'occultations dans le monde du rayonnement : quand une anatomie se superpose à une autre, l'atténuation globale est le produit des atténuations des anatomies. Nous avons affaire à des images **transparentes multiplicativement**.



FIG. 1.5: Exemple d'image obtenue lors d'un examen de graphie. On y observe les coronaires injectées sur le coeur (masse sombre occupant les deux tiers de l'image). Les poumons apparaissent en clair en haut à droite. Les côtes se superposent au tout par transparence. Nous observons également le cathéter injecté de produit de contraste en haut à gauche, ainsi qu'une électrode. L'image est présentée telle qu'obtenue en fin de chaîne image : elle a en particulier été encodée (voir sous-section 1.2.2), et sa dynamique a été adaptée à l'affichage.

1.2.1.2 Propriétés du rayonnement

En pratique, le rayonnement X est obtenu par bombardement d'une anode avec les électrons arrachés à la cathode (chauffées par un filament) et accélérés par la haute tension de l'anode. L'anode réémet l'énergie des électrons qui la heurtent de manière isotrope sous

forme de rayons X. L'ensemble anode-cathode est enserré dans une gaine de plomb dotée d'une fenêtre qui ne laisse sortir le rayonnement X que dans la direction qui nous intéresse. Le principe de fonctionnement du tube est résumé à la figure 1.6.

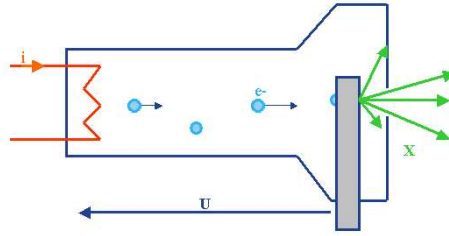


FIG. 1.6: Schéma illustrant le fonctionnement du tube. Le filament (en rouge) est chauffé par un courant intense et émet des électrons dans la chambre vide. Ceux-ci (en bleu) sont accélérés par la tension régnant entre l'anode et la cathode. En heurtant cette dernière (en gris), ils donnent lieu à l'émission de rayons X (en vert). Seuls les rayons partant dans la direction souhaitée ne sont pas arrêtés par la gaine de plomb.

Pour des raisons dues à la nature quantique des phénomènes intervenant lors de l'émission et la transmission des rayons X, il serait inexact de considérer que les photons X suivent exactement les trajectoires rectilignes prédites par la mécanique classique. Ils suivent des lois statistiques, qui peuvent être modélisées en ajoutant un bruit dit **bruit quantique** aux projections calculées par la mécanique classique.

Quand la mécanique classique prédit qu'un flux de N photons X traversera la surface S , il faut tirer le vrai nombre de photons n traversant S suivant la *loi de Poisson* de paramètre N :

$$n \sim \mathcal{P}(N) \quad (1.3)$$

Cette loi a un écart-type \sqrt{N} . Son effet devient complètement négligeable quand le flux est important (comme par exemple dans le cas de la lumière visible), mais il est nettement perceptible quand on travaille, comme dans le cas des images par rayons X, en flux très limité.

Arrivant dans l'anatomie, le rayonnement est donc atténué comme expliqué plus haut. Un autre effet doit être pris en compte : une partie du rayonnement absorbé est réémis de manière isotrope. Cela se traduit par la formation d'un voile flou, le **diffusé**, qui s'ajoute au rayonnement atténué direct qui nous intéresse (Fig.1.7). Nous pouvons modéliser le diffusé par un passe-bas de très basse fréquence \mathcal{D} appliqué sur le rayonnement idéal sortant de l'anatomie $\Phi_{out,ideal}$.

Le diffusé est un rayonnement parasite très important : son ordre de grandeur est supérieur au rayonnement utile. Des grilles, présentées plus loin, permettent cependant de limiter son impact. Insistons sur le fait qu'il introduit une sorte de *transparence additive* entre l'image utile et l'image parasite, et non pas une transparence multiplicative comme l'anatomie. Finalement, le flux de rayons X effectivement mesuré Φ_{out} suit la loi suivante :

$$\Phi_{out} \propto \mathcal{P}\left(\Phi_{out,ideal} + \mathcal{D}(\Phi_{out,ideal})\right) \quad (1.4)$$

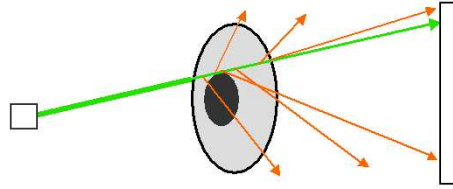


FIG. 1.7: Schéma illustrant le principe du diffusé. En traversant l'anatomie, le rayonnement (vert) est atténué, mais donne aussi lieu à des réémissions partielles, dont certaines aboutissent sur l'écran récepteur.

1.2.2 Mesure des images et prétraitements

Ce flux était jusqu'à quelques années de cela mesuré de manière analogique : les photons X arrivaient sur un scintillateur qui les transformait en de nombreux photons lumineux. Un amplificateur multipliait ensuite ces photons pour qu'ils donnent naissance à une image perceptible, qui était alors filmée par une caméra classique.

Depuis l'an 2000 environ, les constructeurs proposent des technologies numériques bien plus efficaces. Les photons X tombent sur un scintillateur fin qui les convertit en photons lumineux. Ils arrivent alors directement sur une matrice de photodiodes, qui les convertit en signal électrique, par la suite converti en signal numérique (Fig.1.9). Cette conversion est entachée d'un **bruit électronique** faible, mais pas toujours négligeable aux bas niveaux de gris.

En outre, la conversion photons X - photons lumineux introduit un autre biais dans la formation de l'image : un photon X ne donne pas naissance seulement à des photons lumineux qui suivent exactement la direction d'incidence. Certains d'entre eux seront mesurés par des cellules voisines de celle qu'aurait heurté le photon X directement.

Cet effet peut être modélisé dans une **fonction de transfert du détecteur** (ou **MTF** pour *Modulation Transfer Function*). Nous y intégrons aussi les déformations du signal dues à la taille finie des diodes. L'action de la MTF peut être vue mathématiquement comme la convolution d'une image "idéale" avec un noyau 3×3 ayant un net pic en 0. L'étalement du spectre est en effet fortement limité par la structure en aiguilles parallèles du scintillateur. Ce filtre est représenté à la Fig.1.8.

L'écran, composé donc du scintillateur et des diodes, est surmonté d'une *grille anti-diffusion*. Elle est composée de rangées horizontales et verticales de petites lamelles en plomb perpendiculaires à l'écran. Celles-ci arrêtent les photons X trop inclinés par rapport à l'écran, et qui ne pouvaient donc pas venir de la source d'émission (Fig.1.9).

En fait, elles arrêtent aussi une partie du rayonnement primaire, mais permettent de fortement réduire l'influence relative du diffusé : il ne forme plus que 20 à 50% du rayonnement mesuré sur l'écran.

Les charges électroniques mesurées sur le récepteur sont alors converties en niveaux de gris, et codés sur 14 bits. L'image électronique obtenue est encore difficile à appréhender en l'état : elle représente une image des atténuations, alors que nous comprenons plus facilement les images d'*épaisseur radiologique*. Pour passer du "monde" des $\prod e^{-\mu_i \cdot d_i}$ à celui

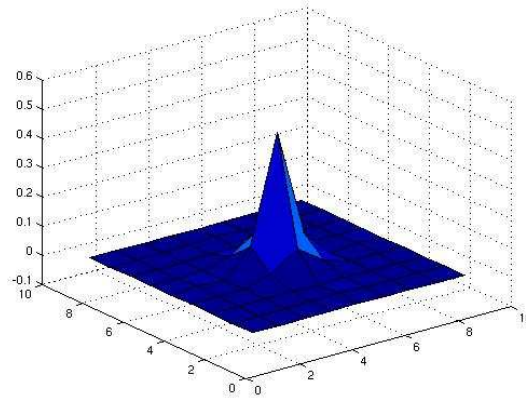


FIG. 1.8: Allure de la MTF du détecteur. Il s'agit d'un noyau 9×9 , très proche d'un noyau 3×3 . Le pic central est 5 fois plus haut que ses voisins directs, et 15 fois plus haut que ses voisins plus éloignés.

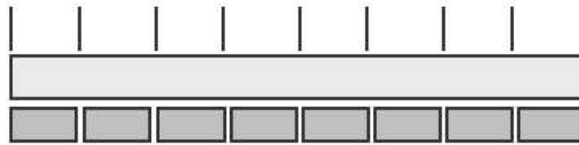


FIG. 1.9: Schéma du détecteur. De haut en bas : la grille anti-diffusion qui arrête les rayonnements trop obliques, le scintillateur qui convertit les photons X en photons lumineux, puis les photodiodes mesurant les photons lumineux sous forme de charge électrique.

des $\sum \mu_i \cdot d_i$, nous introduisons une transformation par un opérateur d'**encodage**².

L'opérateur le plus naturel serait le logarithme log. On observe expérimentalement que le diffusé ne perturbe pas la perception visuelle de l'image après transformation, sa structure très basse fréquence qui s'apparentant localement à un offset.

D'autre part, comme nous préférons travailler sur des images entachées d'un bruit gaussien plutôt que poissonnien, l'opérateur racine carrée serait intéressant. La transformation de Anscombe [Anscombe 48] permet en effet de transformer le bruit poissonnien en bruit gaussien (en première approximation). Elle ne pose problème qu'aux bas niveaux de gris car alors le bruit électronique prend le pas sur le bruit quantique, et la racine carrée peut l'amplifier à tort.

En fait, nous montrons que les différents algorithmes proposés par la suite se comportent

²Nous employons pour désigner cet opérateur, qui correspond en fait à une transformation suivie d'une quantification sur un nombre donné de bits, le terme "d'encodage" utilisé en interne à General Electric Healthcare. Historiquement, il avait en effet outre la fonction de transfert du "monde" du rayonnement à celui des épaisseurs un rôle de réduction du volume des données. Il transformait des images codées en 14 bits à des images codées en 12 bits, car il n'était alors pas possible de traiter en temps réel des images 14 bits.

de manière similaire que l'on choisisse la transformation par le logarithme ou celle par la racine carrée. Ceci s'explique par le fait que les fonctions en question se ressemblent beaucoup sur l'intervalle correspondant à la dynamique de l'image (Fig.1.10). Nous travaillerons alors, sauf mention contraire, avec une fonction racine carrée, prolongée linéairement aux bas niveaux pour ne pas trop amplifier le bruit électronique.

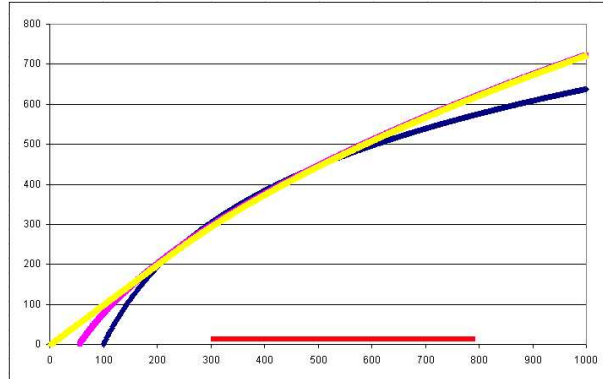


FIG. 1.10: Trois possibilités de fonctions d'encodage : log en bleu, racine carrée en rose et racine carrée prolongée de manière linéaire aux bas niveaux en jaune. Ces transformations ont été étalonnées pour être cohérentes avec la dynamique de l'image. Nous notons une grande similarité entre elles dans la zone où l'écrasante majorité des niveaux de gris se situe : entre 300 et 800 niveaux de gris (le niveau moyen des images est fixé à 512 sur ces images 14 bits).

Nous résumons la physique de formation des images par rayons X sur la figure 1.11.

1.2.3 Caractéristiques des images et objectifs de la thèse

Ce système d'imagerie permet de visualiser en temps réel l'anatomie du patient, et ouvre donc la voie aux méthodes de diagnostic et de traitement de maladies vasculaires présentées à la section 1.1. Cependant, l'utilisation de rayonnements X n'est jamais bénigne, car elle favorise la naissance de cancers, à la fois chez le patient et chez le personnel médical. L'exposition est donc limitée autant que possible.

Les examens à des fins de diagnostics étant courts (ordre de grandeur d'une minute d'exposition maximum), on peut autoriser des rayonnements assez importants. Ce mode de fonctionnement, la **graphie**, donne naissance à des images au RSB (*rapport signal sur bruit*) relativement fort (de l'ordre de 50).

Par contre, le mode choisi pour les opérations de traitement, la **fluoroscopie**, utilise des doses 5 à 10 fois plus faibles. Une opération délicate peut en effet donner lieu à des dizaines de minutes d'exposition. Les images sont alors beaucoup plus bruitées, avec des RSB de 15 à 25³.

D'autre part, les zones imagées sont relativement uniformes : les tissus humains apparaissent assez peu texturés aux rayons X. Les structures d'intérêt (les outils interventionnels

³Le RSB, qui peut sembler élevé, rend peu compte de la difficulté des images. Un indicateur plus éloquent est le RCB (Rapport Contraste sur Bruit) des objets d'intérêt, qui se situe dans la plage [0.5,3].

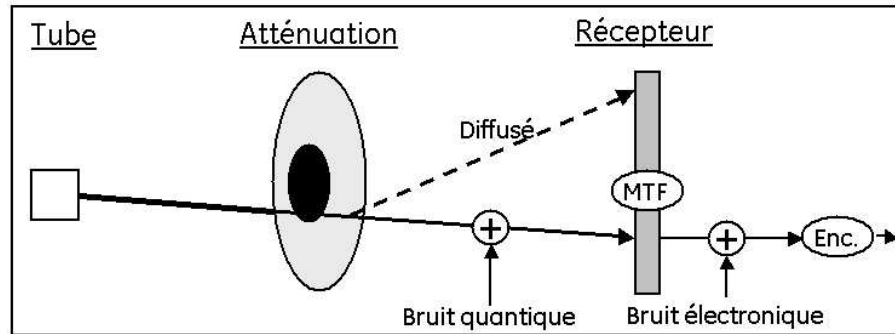


FIG. 1.11: Schéma résumant la physique de formation des images. Le rayonnement traverse l'anatomie et y est atténué. Ce faisant, un diffusé est créé et reste présent malgré une réduction considérable par la grille anti-diffusé. Le rayonnement entraîne un bruit poissonnier, auquel s'ajoute le bruit électronique du détecteur. Le comportement du scintillateur et des diodes est modélisé par une MTF connue. Enfin, l'image mesurée passe par un opérateur d'encodage pour être interprétable en tant qu'épaisseur.

essentiellement) étant en outre peu contrastées, les images fluoroscopiques apparaissent comme relativement uniformes, peu contrastées et donc très bruitées (relativement à l'information intéressante dans les images).

Il est crucial de réduire le niveau de bruit de ces images, à la fois pour permettre au praticien de bien distinguer les objets d'intérêt, et aussi pour limiter sa fatigue visuelle après des heures de concentration sur ses écrans.

Nous présentons des images des deux types à la figure 1.12.

C'est dans ce cadre que se situe cette thèse. Nous voulons proposer des algorithmes de réduction de bruit pour les images fluoroscopiques, et plus particulièrement les images cardiaques car ce sont celles qui ont le niveau de bruit le plus fort (puisque les rayons doivent traverser tout le thorax, souvent dans des angulations difficiles).

Il y a bien sûr un très vaste champ de méthodes possibles pour ce faire, que nous évoquerons dans le chapitre 6. Toutes les approches de débruitage intégrant une dimension spatiale (filtres adaptatifs spatio-temporels, filtres dans l'espace de Fourier 3D, filtres dans des espaces d'ondelettes...) posent le problème de la coloration du bruit. Ils permettent d'obtenir des augmentations de RSB impressionnantes, mais produisent des artefacts très gênants de regroupement de bruit en grains, de sorte que le bruit perçu par un observateur semble au contraire augmenté. Nous sommes d'autant plus sensibles à ce genre de phénomènes que, pour appréhender les rapides mouvements cardiaques, nous travaillons à une cadence d'acquisition élevée (30 images par seconde). Dans ces conditions, il est très délicat de régler les algorithmes de débruitage spatial de sorte à ce que leur performance ressentie approche leur performance théorique.

Alors que d'autres travaux cherchent à contrôler ces effets, nous avons au contraire choisi de nous placer dans cette thèse essentiellement dans le cadre du **débruitage temporel pur**. Il a en effet l'avantage de n'introduire aucun artefact de coloration de bruit, et permet de conserver l'allure générale de l'image à laquelle les médecins sont très attachés. Il se

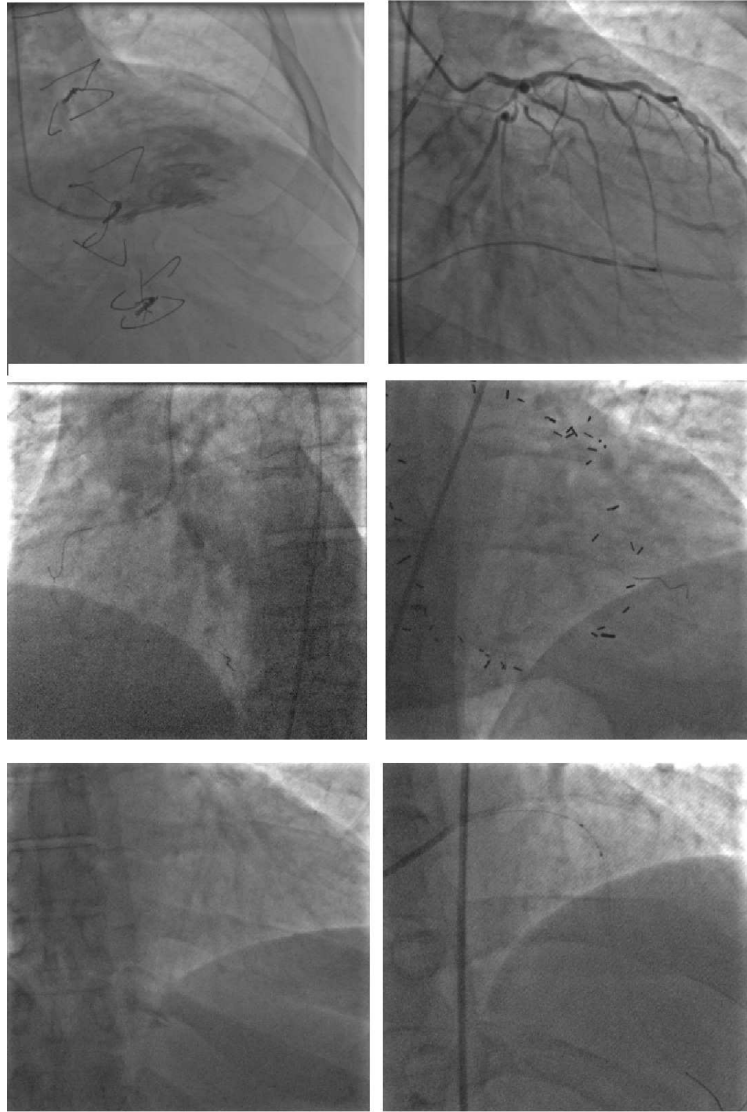


FIG. 1.12: Diverses anatomies imagées selon différents modes. Rangée du haut : graphie. À gauche, ventriculographie : un cathéter injecte du produit de contraste dans un ventricule pour évaluer son volume. Le coeur forme la masse sombre principale, les poumons sont visibles en haut à droite et les côtes se superposent au tout. Des agrafes chirurgicales sont visibles. À droite, coronographie. Les artères coronaires sont injectées, sur le coeur. Des électrodes sont également visibles dans les oreillettes. Rangée centrale : des fluoroscopies, donc plus bruitées. À gauche, le coeur, centré sur l'image, se superpose au diaphragme (en bas à gauche), aux poumons (en haut à gauche) et à la colonne vertébrale (à droite). Deux guides sortent d'un cathéter. A droite, le coeur est imagé de plus près. On observe un cathéter, des agrafes, et un guide, principalement par son extrémité plus opaque. Rangée du bas, deux fluoroscopies. À gauche, le coeur imagé dans une angulation qui le superpose au diaphragme, sans outil dans l'image. À droite, une intervention cardiaque avec pose de stent entre les deux marqueurs au centre de l'image.

prête en outre bien à une implantation temps réel. Notre contribution majeure porte sur la compensation de mouvement en présence de transparence, qui permet d'utiliser un filtrage temporel significatif sans faire disparaître de détails.

Notre thèse porte donc sur le **débruitage temporel de séquences fluoroscopiques par filtrage dans le sens du mouvement**. Comme les méthodes de débruitage spatio-temporel profitent aussi de la compensation du mouvement, nous étudions également dans le chapitre 9 un filtre spatio-temporel compensé en mouvement. Cette étude est à prendre comme une extension de notre approche initiale, dont la philosophie consiste à n'utiliser que la dimension temporelle pour réduire le bruit afin de ne pas modifier l'esthétique des images.

1.3 Des images transparentes

Comme nous l'avons déjà souligné, les images que nous considérons ont la propriété d'être **transparentes**. C'est en cela que notre travail est original vis à vis des nombreuses contributions sur la compensation de mouvement : la physique régissant la formation des images est toute autre, et les méthodes d'estimation comme de compensation de mouvement doivent donc être repensées.

Nous détaillons dans ce paragraphe le modèle des couches que nous reprenons pour formaliser le problème. Nous introduisons ensuite une méthode de simulation d'images cliniques réalistes, à mouvement connu, qui nous permettra d'évaluer les algorithmes proposés dans cette thèse.

1.3.1 Modèle des couches

La physique des rayons X donne naissance à des images transparentes multiplicativement. Un encodage permet de se ramener dans le monde des épaisseurs radiologiques, où la transparence est additive.

Nous nous représentons alors l'image mesurée I comme une superposition de différentes sous-images I_i , animées par leur mouvement propre. En chaque pixel \mathbf{p} et à la date t , nous avons alors :

$$I(\mathbf{p}, t) = \sum_i I_i(\mathbf{p}, t) \quad (1.5)$$

Chacune de ces couches vérifie de son côté la contrainte de conservation de l'intensité. Cela revient à décomposer l'image mesurée en une superposition de couches qui ne se replient pas sur elles-mêmes. Notant $\mathbf{w}_i(\mathbf{p}, t)$ le déplacement rétrograde de la couche i au point \mathbf{p} entre les dates t et $t - 1$, l'équation suivante est vérifiée :

$$I(\mathbf{p}, t) = \sum_i I_i(\mathbf{p} + \mathbf{w}_i(\mathbf{p}, t), t - 1) \quad (1.6)$$

Dans le cadre de cette formalisation du problème, notre objectif est de concevoir des méthodes pour estimer les mouvements $\mathbf{w}_i(\mathbf{p}, t)$, et de les utiliser afin de débruiter temporellement la séquence I .

Attardons-nous sur cette notion de couches. Intuitivement, nous voudrions considérer les organes comme différentes couches : la couche correspondant au coeur, la couche

des poumon, la couche du diaphragme, celle de colonne vertébrale, celle des outils interventionnels, etc. Il serait néanmoins plus intéressant de considérer comme une seule et même couche des organes ayant un mouvement cohérent. Ainsi, le coeur entraîne dans son mouvement des tissus qu'il serait opportun d'attribuer à la même couche.

À l'inverse, sous certaines angulations, les deux parois du coeur apparaissent superposées avec des mouvements distincts. Il faudra alors les considérer comme deux couches distinctes.

En fait, il faut comprendre que le critère de constitution des couches est la **cohérence des mouvements** :

Définition 1 *Nous considérons qu'une **couche** est un ensemble physique animé d'un mouvement cohérent, sous l'angulation d'acquisition.*

Cette notion recoupe d'ailleurs très largement la segmentation en organes, puisque ce sont eux qui impriment les différents mouvements présents à l'image.

La définition 1 est satisfaisante dans notre optique, d'autant plus qu'elle est cohérente avec l'intuition de l'observateur face à une séquence : il perçoit comme couches distinctes les tissus se superposant et bougeant différemment.

La notion de couche reste cependant sujette à la définition du *mouvement cohérent*. Comme nous le verrons plus loin, cela s'intègre bien dans notre approche car nous modéliserons les mouvements des couches sous forme paramétrique dans les contributions présentées plus loin. Mais d'autres travaux pourraient proposer d'autres classes de mouvement comme référence.

1.3.2 Simulation d'images réalistes

Pour tester nos algorithmes d'estimation de mouvement et de débruitage, nous voulons générer des images synthétiques réalistes, c'est-à-dire qui suivent la physique de la formation des images X. Ceci nous permettra de contrôler les mouvements et le niveau de bruit des séquences tests, et donc d'évaluer objectivement et quantitativement la performance de nos algorithmes, la vérité terrain étant connue.

Nous procédons en deux temps. Nous commençons par expliquer comment nous créons des cartes d'atténuation typiques des anatomies considérées, puis nous expliquons comment nous les combinons pour générer des séquences d'images réalistes, en nous appuyant sur le formalisme des couches.

1.3.2.1 Génération de cartes d'atténuation de référence

Nous souhaitons dans un premier temps inverser le diagramme de formation des images de la figure 1.11, et donc de reconstruire la carte d'atténuation "physique" à partir d'images cliniques mesurées. Nous choisissons en effet de partir d'images réelles afin de simuler des images ayant la dynamique typique des examens par rayons X pour garder un contenu parfaitement réaliste. Notre objectif est en effet de simuler la production d'images (à mouvements connus), mais en aucun cas leur contenu.

Pour nous affranchir du bruit, nous choisissons des images de graphie de très bonne

qualité où le bruit est faible, et que nous considérons comme non bruitées. Nous discutons de l'impact des différentes approximations plus bas. Cette procédure a l'avantage de nous éviter de chercher à inverser les opérateurs de bruit quantique et de bruit électronique.

Nous commençons alors par inverser l'opérateur d'encodage, ce qui permet de reconstruire l'image d'atténuation sortant du détecteur. Nous inversons ensuite la MTF en convoluant cette image avec le filtre inverse.

À ce point de l'inversion du diagramme de formation des images, nous obtenons l'image d'atténuation idéale en entrée du détecteur. Cependant, elle s'exprime toujours en niveaux de gris. Puisque nous connaissons les rapports *électrons / niveaux de gris*, *photons lumineux / électrons* des diodes et *photons X / photons lumineux* du scintillateur, nous transformons notre image de niveaux de gris en image de rayonnement à l'entrée du détecteur.

Ensuite, pour limiter l'effet du diffusé, nous estimons sa valeur afin de la retirer à l'image. Pour ce faire, nous appliquons un passe-bas très basse fréquence. Nous avons choisi un moyenneur 64×64 en nous basant sur l'expérience des ingénieurs de GE travaillant sur le sujet. Nous ôtons alors à l'image une partie (typiquement 20%), de l'image basse fréquence que nous venons de filtrer.

Ceci fait, nous avons calculé le rayonnement principal sortant du patient. Il ne reste plus qu'à diviser par le rayonnement en entrée pour obtenir l'image d'atténuation souhaitée.

Lors de ce travail de génération de cartes d'atténuations, nous avons fait les approximations suivantes :

- Nous faisons l'hypothèse que les images choisies sont non bruitées. En fait, elles contiennent un léger bruit, qui entache la carte d'atténuation d'un faible bruit spatial. Cela dit, l'utilisation de filtres spatiaux sur des images à la texture si fine dénaturerait l'apparence de l'anatomie. Il est exact que le niveau de bruit des séquences traitées jouera un rôle essentiel dans nos simulations à venir, mais il s'agira d'un bruit temporel, variant le long de la séquence. A l'inverse, nous parlons ici d'un bruit spatial associé à une image unique, et dont la présence fausse seulement légèrement l'apparence de la couche correspondante mais ce, d'une manière identique tout le long de la séquence synthétique générée.
- La MTF étant à support fini, son inverse est à support infini et peut donc seulement être approché. En fait, nous verrons dans le sous-paragraphe suivant que nous pouvons éviter l'étape de l'inversion de la MTF.
- Enfin, la correction du diffusé est clairement empirique. Il faut encore une fois voir qu'il se traduit par un voile flou sur toute l'image, et qu'il n'est pas essentiel de l'annuler exactement, d'autant plus que nous allons recombinaisonner les cartes d'atténuation pour simuler de nouvelles images, et donc simuler un nouveau diffusé, tout aussi flou.

1.3.2.2 Synthèse de séquences réalistes

Nous considérons que ces cartes d'atténuation sont des candidats pertinents pour jouer le rôle de couches dans notre modèle. Elles présentent en effet la dynamique particulière des couches "physiques" puisqu'elles imagent par rayons X les organes qui nous intéressent. Nous sommes alors en mesure de générer des couches, dans le monde des atténuations, représentatives de différentes parties de l'anatomie : coeur, abdomen, colonne vertébrale,

ou même cérébral ou membres inférieurs. Trois exemples de choix de couches possibles sont présentés à la figure 1.13. Nous pouvons leur imprimer les mouvements qui nous intéressent et les combiner entre elles : une superposition se traduira par une *multiplication* des atténuations.

Ce faisant, nous définissons une séquence d'atténuations dont nous connaissons la

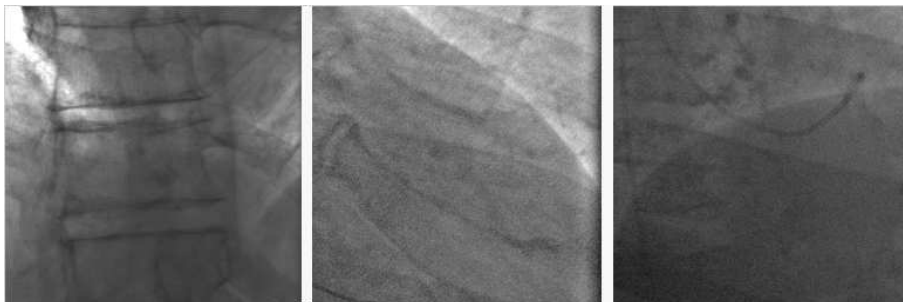


FIG. 1.13: Trois exemples de choix de couches possibles (images 512×512). À gauche, une vue abdominale. Au centre, une vue cardiaque. À droite, une vue cardiaque avec le diaphragme, et une électrode.

configuration et les vitesses. Il reste alors à simuler une acquisition par le système de visualisation de cette séquence, en suivant le principe illustré à la figure 1.11.

Nous choisissons le rayonnement en entrée, qui pourra être typique d'une graphie ou d'une fluoroscopie. Nous ajoutons le diffusé, calculé comme précédemment, en ajoutant 20 à 50% d'un filtrage de très basse fréquence de l'image de rayonnement. Nous simulons alors le bruit quantique, puis transformons cette image dans l'espace des niveaux de gris. Pour simuler toute l'action du détecteur, nous convoluons l'image avec la MTF et lui ajoutons un bruit électronique. Enfin, l'image est soumise à l'opérateur d'encodage.

Notons que, du fait de la linéarité de la MTF, il n'est pas nécessaire de l'inverser quand on construit les cartes d'atténuation et de la réappliquer lors de la simulation du rayonnement. C'est pourquoi nous n'avons pas besoin de l'inverser en construisant les cartes, à condition de prendre garde à l'appliquer au bruit quantique simulé.

Nous proposons donc finalement une méthode de génération d'images cliniques réalistes à vérité terrain contrôlée. Nous pouvons choisir :

- le nombre et le type de couches, ainsi que leur configuration.
- leur mouvement.
- leur niveau de bruit via le niveau de radiation appliquée.
- le niveau de diffusé.

Ces images artificielles ont été construites à partir d'images réelles pour avoir une dynamique typique des anatomies imagées, et elles suivent la physique de l'imagerie par rayons X. Le processus de construction de ces images à partir des cartes d'atténuation est résumé en figure 1.14.

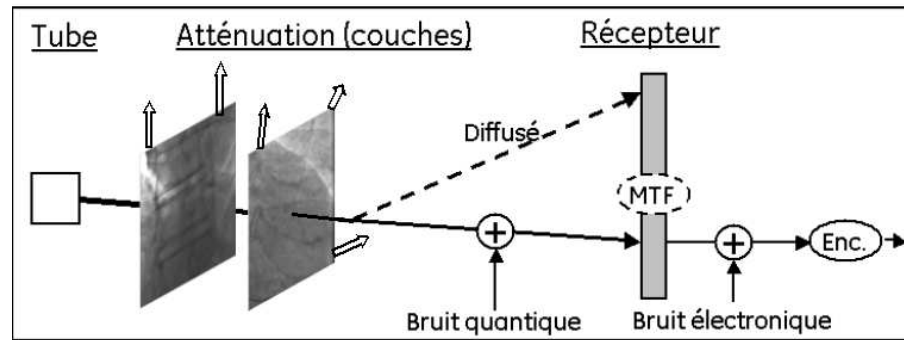


FIG. 1.14: Schéma résumant la génération des images-test réalistes. Nous plaçons les couches d'atténuation dans la configuration qui nous intéresse, en leur imprimant un mouvement connu. Puis, nous simulons la chaîne d'acquisition : rayonnement, diffusé, bruit quantique, conversion en niveaux de gris, bruit électronique, MTF (sur l'image ou juste sur le bruit quantique) et encodage.

1.4 Conclusion

Nous avons présenté dans cette partie le cadre clinique dans lequel s'inscrit la modalité d'imagerie par rayons X qui nous intéresse dans cette thèse. L'angiographie et la cardiographie permettent un diagnostic des maladies vasculaires en imageant le réseau artériel du patient, et mettent en particulier à jour d'éventuelles sténoses. Une des possibilités de traitement est l'angioplastie : le praticien intervient mécaniquement pour traiter les sténoses en guidant son geste grâce au système d'imagerie par rayons X.

Ces images sont donc essentielles à un bon diagnostic et un cathétérisme précis. Comme elles sont acquises sous rayonnement aussi limité que possible pour préserver la santé du patient et du personnel médical, les images obtenues contiennent un bruit important, surtout quand on le compare avec le contraste des structures d'intérêt.

Notre thèse vise à concevoir des algorithmes de débruitage. Nous avons pris le parti de nous focaliser sur le débruitage temporel pur, sans composante spatiale (à part dans le chapitre 9), car il garantit une conservation de l'apparence de l'image à laquelle le médecin est habitué, et évite les problèmes de coloration du bruit qui ont un impact important aux hauts taux d'acquisition utilisés.

L'originalité de ce travail est de considérer des images dont la physique de formation est différente des images vidéo classiques. Elles sont entachées par un fort bruit quantique (coloré spatialement par la MTF), et par le diffusé qui forme un voile basse fréquence recouvrant l'image. Surtout, puisqu'il s'agit d'images obtenues par transmission et non par réflexion, elles présentent une transparence additive (grâce à l'encodage de la figure 1.10). Quand un objet se trouve devant un autre, il n'y a pas d'occultation mais leurs niveaux de gris s'ajoutent.

Cette notion est formalisée en introduisant le concept de couche, définie au sens du mouvement. Nous avons finalement décrit une méthode de génération d'images test réalistes, sur lesquelles nos algorithmes pourront être testés et évalués à vérité-terrain connue.

Deuxième partie

Estimation de mouvements
transparents

Chapitre 2

Estimation du mouvement avec transparence ; état de l'art

Notre travail porte donc sur le développement d'algorithmes de réduction de bruit par filtrage temporel sur des séquences d'examen obtenues par rayons X. Sa grande originalité est de porter sur des images transparentes. En effet, la physique même des rayons X induit une transparence multiplicative (comme expliqué dans le chapitre 1). Un opérateur d'encodage la transforme ensuite en **transparence additive**.

Nous commençons cette deuxième partie par un état de l'art de l'estimation de mouvement transparent. Ce sujet a donné lieu à de multiples contributions, plus souvent inspirées par les phénomènes de transparence apparaissant dans les séquences vidéo classiques que par l'imagerie médicale. Il peut s'agir par exemple de réflexions dans l'eau, un miroir ou un cadre. Ce phénomène est également présent au cinéma dans certains effets spéciaux, par exemple ceux qui représentent les fantômes par des êtres transparents. Citons encore les fondus enchaînés qui permettent une transition progressive d'une scène à l'autre (Fig.2.1). Nous présentons pour commencer les premiers travaux ayant abordé ce problème, dans des cas particuliers qui ne nécessitaient pas le développement d'un formalisme intégrant la nature transparente des images. Il est devenu clair au fil de ces contributions que le problème ne pourrait être résolu de manière complètement satisfaisante qu'en modélisant explicitement la transparence. C'est à ce titre que les travaux de Shizawa et Mase ont marqué un tournant, même si leur utilisation pratique est restée limitée.

Mais ils ont ouvert la voie aux deux grandes classes d'approches permettant d'estimer de manière fiable les mouvements par transparence : l'approche fréquentielle qui cherche à séparer des plans dans l'espace de Fourier 3D, et l'approche directe qui se base sur un équivalent à l'équation de conservation de l'intensité. Une section est consacrée à chacune de ces deux méthodes.

Enfin, nous évoquerons dans le dernier paragraphe une question connexe, celle de la séparation des couches. Il s'agit alors non pas d'estimer les vitesses des couches transparentes présentes, mais de les séparer à partir de la seule donnée de la séquence transparente.

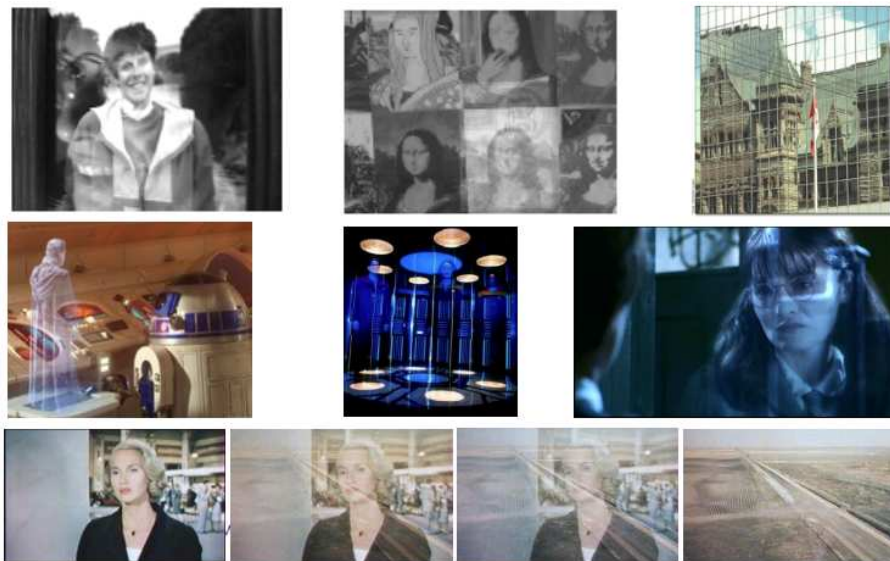


FIG. 2.1: Différentes situations de transparence dans des images vidéo. En haut, trois situations de réflexion. À gauche, une personne se reflète dans le cadre d'une photographie. Au centre, un sachet de corn-flakes se reflète dans une vitre couvrant un portrait statique. À droite, un bâtiment se reflète dans un gratte-ciel. Au centre en hauteur, trois exemples d'effets spéciaux. À gauche, un hologramme. Au centre, une téléportation. À droite, un fantôme. En bas, un exemple de fondu enchaîné.

2.1 La problématique de la transparence : vers des estimateurs dédiés

2.1.1 Adaptation partielle d'estimateurs de mouvements classiques

Il arrive que la transparence apparaissant dans les séquences vidéo puisse être traitée par des méthodes ne la modélisant pas explicitement. Le phénomène peut en effet être marginal et dévier assez peu des hypothèses classiques de conservation de l'intensité, de sorte que des méthodes d'estimation de mouvement suffisamment robustes arrivent à calculer les mouvements présents de manière satisfaisante.

Seulement, plus les spécificités des images transparentes sont intégrées au mécanisme d'estimation, plus les méthodes sont applicables à de larges classes d'images transparentes. Il devient finalement clair que seule une modélisation complète du phénomène pourra prétendre à estimer de manière générale les mouvements transparents.

Black et Anadan [Black 96a] cherchent à profiter de la robustesse de leurs estimateurs pour traiter des situations de transparence. Alors que les estimateurs classiques supposent qu'il n'y a qu'un seul mouvement localement, les estimateurs robustes sont peu sensibles à cette hypothèse.

Les auteurs utilisent une méthode d'estimation itérative des mouvements dominants : ils estiment par régression affine robuste le mouvement dominant, puis recommencent sur les

outliers à ce mouvement s'ils sont assez nombreux. Leur article en démontre la robustesse en exhibant des estimées convaincantes sur des situations mettant à mal l'hypothèse d'unicité du mouvement localement.

Les exemples donnés concernent des problèmes de discontinuité de profondeur, d'occlusion fragmentée¹ ou des cas particuliers de transparence. Ils sont révélateurs du fait que les auteurs cherchent à démontrer la solidité des estimateurs robustes sur des images difficiles, plus qu'à proposer une méthode d'estimation du mouvement transparent valide dans le cas général.

Quoiqu'il en soit, les exemples donnés présentent tous un très net mouvement dominant ; dans le cas de la transparence, une couche est beaucoup plus contrastée que l'autre, de sorte qu'un estimateur robuste peut extraire de l'image le mouvement dominant (il s'agit de l'image en haut à gauche de la figure 2.1). Les auteurs eux-mêmes soulignent que leur approche ne permet pas de traiter la transparence de manière générale.

Irani et Peleg proposent dans [Irani 93] et [Irani 94] une approche alternative aux estimateurs robustes pour l'estimation du mouvement de plusieurs objets. Ils estiment dans un premier temps la translation dominante de l'image (car les translations globales sont robustes aux petits mouvements parasites). Ceci leur permet de segmenter la zone dominante, sur laquelle ils estiment ensuite un mouvement plus complexe (affine, projectif, etc). Ils recommencent alors le même travail sur le reste de la scène.

L'article présente une évolution de cette technique pour le suivi d'un objet segmenté *le long d'une séquence* : on estime alors le mouvement non pas entre l'image passée et l'image actuelle, mais entre la moyenne de toutes les images précédentes, recalées sur l'objet d'intérêt, et l'image actuelle. Ceci permet de flouter le fond (qui perturbera peu l'estimation), et de débruiter l'objet d'intérêt.

Cette méthode est étendue aux mouvements transparents. A la différence de [Black 96a] qui appliquait directement son estimateur robuste aux images transparentes, les auteurs font un premier pas vers une modélisation de la transparence.

Dans le cas où une des deux couches est nettement plus contrastée que l'autre, la méthode proposée jusque là permet d'estimer son mouvement. Une moyenne des images recalées permet de reconstruire la couche dominante correspondante. Un masque segmentant les pixels faisant partie de la deuxième couche est alors construit en prenant en compte la transparence : il s'agit des pixels où l'image mesurée n'est pas expliquée de manière satisfaisante par la couche dominante. L'estimation est alors relancée sur ce second masque.

Là aussi, nous avons affaire à une approche qui suppose l'existence d'une couche dominante, ce qui permet de se contenter d'estimations successives. La méthode proposée par les auteurs n'est en fait valide que pour des mouvements qui se dominent mutuellement de manière nette, et qui sont proches de translations.

Bergen *et al.* [Bergen 92b] partent d'un modèle de transparence exact : en tout pixel \mathbf{p} et à toute date t , l'image I est la somme des deux couches I_1 et I_2 qui se déplacent aux vitesses $\mathbf{w}_1 = (u_1, v_1)$ et $\mathbf{w}_2 = (u_2, v_2)$. Nous supposons les mouvements constants

¹Cette dernière configuration se rapporte aux cas où un objet en mouvement passe derrière un objet fragmenté type haie ou barrière. Leurs supports sont intimement mêlés mais la conservation de l'intensité est vérifiée.

spatialement et temporellement aux trois dates $t - 1$, t et $t + 1$.

$$I(\mathbf{p}, t) = I_1(\mathbf{p}, t) + I_2(\mathbf{p}, t) \quad (2.1)$$

$$I_1(\mathbf{p}, t + 1) = I_1(\mathbf{p} + \mathbf{w}_1, t) \quad (2.2)$$

$$I_2(\mathbf{p}, t + 1) = I_2(\mathbf{p} + \mathbf{w}_2, t). \quad (2.3)$$

Supposons la vitesse \mathbf{w}_1 connue. Nous pouvons alors former :

$$\begin{aligned} I(\mathbf{p}, t + 1) - I(\mathbf{p} + \mathbf{w}_1, t) &= I_1(\mathbf{p}, t + 1) + I_2(\mathbf{p}, t + 1) \\ &\quad - I_1(\mathbf{p} + \mathbf{w}_1, t) - I_2(\mathbf{p} + \mathbf{w}_1, t) \end{aligned} \quad (2.4)$$

$$\begin{aligned} &= (I_1(\mathbf{p}, t + 1) - I_1(\mathbf{p} + \mathbf{w}_1, t)) \\ &\quad + (I_2(\mathbf{p}, t + 1) - I_2(\mathbf{p} + \mathbf{w}_1, t)) \end{aligned} \quad (2.5)$$

$$= I_2(\mathbf{p}, t + 1) - I_2(\mathbf{p} + \mathbf{w}_1, t) \quad (2.6)$$

en reconnaissant (2.2). De la même façon,

$$I(\mathbf{p}, t) - I(\mathbf{p} + \mathbf{w}_1, t - 1) = I_2(\mathbf{p}, t) - I_2(\mathbf{p} + \mathbf{w}_1, t - 1) \quad (2.7)$$

$$= I_2(\mathbf{p} - \mathbf{w}_2, t + 1) - I_2(\mathbf{p} + \mathbf{w}_1 - \mathbf{w}_2, t - 1). \quad (2.8)$$

On reconnaît (2.6), décalée de $-\mathbf{w}_2$. Autrement dit, si la vitesse d'une couche est connue ($-\mathbf{w}_1$ par exemple), nous pouvons, en formant les différences recalées (2.6) et (2.8), construire deux images *sans transparence* décalées de $-\mathbf{w}_2$. Une estimation monocouche classique appliquée à ces deux cartes de différences permet donc de trouver la vitesse manquante $-\mathbf{w}_2$.

Les auteurs proposent alors un schéma d'estimation alterné d'une vitesse, puis de l'autre. La convergence d'un tel algorithme est pourtant hautement improbable, sauf si on suppose implicitement qu'une couche domine clairement l'autre et que son mouvement peut être trouvé directement (ce qui est le cas sur l'exemple donné).

Szeliski, Aridan et Anadan dans [Szeliski 00] proposent une approche médiane entre utilisation d'estimateurs classiques et modélisation de la transparence. Ils se concentrent sur la transparence à deux couches, et cherchent à estimer de manière jointe leur mouvement et leurs niveaux de gris. Si le coeur de leur travail repose bien sur une modélisation explicite de la transparence, nous verrons qu'il est dépendant d'une initialisation basée sur des estimateurs robustes classiques.

Les auteurs proposent une méthode de minimisation alternée :

- À mouvements de couches connus, ils approchent le niveau de gris d'une des deux couches par la valeur minimale de chaque pixel le long de la séquence recalée. En effet, une fois le mouvement de la couche compensé, on observe en chaque point la valeur de la couche immobile, à laquelle s'ajoute celle de la couche en mouvement. La plus petite des valeurs observées sur chaque pixel de la séquence recalée peut donc fournir une première estimation intéressante de la couche fixe.

On peut alors approcher la seconde couche en soustrayant la première, puis en prenant le maximum de cette nouvelle séquence, recalée cette fois par le second mouvement. Les auteurs montrent qu'on finit par obtenir, dans des conditions idéales, des estimés corrects des deux couches par itérations de cette méthode.

- Une fois les deux couches extraites, on peut estimer leur mouvement en minimisant une énergie.

Le mécanisme d'estimation des couches à mouvement connu fait des hypothèses fortes (les “*conditions idéales*”). Il suppose qu'un nombre suffisant de pixels de chaque couche vaut 0, ce qui permet peu à peu de remonter à la vraie valeur de chaque couche. En outre, les interpolations lors des recalages sont problématiques. Et surtout, les auteurs insistent sur le fait que la reconstruction des couches ne supporte pas le moindre bruit sur la séquence. Mais la plus grosse entorse à la transparence vient du fait que son hypothèse est relaxée au début de l'algorithme. Une méthode itérative de ce type suppose en effet une bonne initialisation. Les auteurs proposent d'estimer les mouvements dans un premier temps de manière robuste par la méthode de [Black 96a] présentée plus haut. Les limitations de cette technique rejaillissent donc sur l'estimateur de [Szeliski 00] : dans la pratique, une couche doit être nettement plus contrastée que l'autre pour que cela fonctionne.

Ces premiers travaux montrent qu'il est impossible d'estimer de manière générale les mouvements transparents sans utiliser de modèle exact de la physique des images en présence. Même si dans certains cas particuliers intéressants des méthodes robustes, en partie adaptées à la transparence, estiment correctement les mouvements transparents, elles ne peuvent pas prétendre traiter des situations de transparence très générales comme celles présentes dans les images d'examen par rayons X.

2.1.2 Une première modélisation complète de la transparence par généralisation du flot optique

2.1.2.1 Principe

Ce sont Shizawa et Mase qui les premiers ont proposé un formalisme complet modélisant explicitement la transparence. Ils adoptent dans leurs trois articles ([Shizawa 90], [Shizawa 91b] et [Shizawa 91a]) une approche mathématique qui ouvre la voie aux deux grands types de méthodes d'estimation de mouvement transparent : l'approche fréquentielle et l'approche directe, détaillées dans les deux sections suivantes.

Les auteurs introduisent l'opérateur de flot optique dans ces deux espaces :

$$\mathbf{a}(\mathbf{w}) = \begin{bmatrix} u & v & -1 \end{bmatrix} \cdot \begin{bmatrix} \partial_x \\ \partial_y \\ \partial_t \end{bmatrix} \text{ dans l'espace direct,} \quad (2.9)$$

$$\tilde{\mathbf{a}}(\mathbf{w}) = 2\pi \cdot \mathbf{j} \cdot \begin{bmatrix} u & v & -1 \end{bmatrix} \cdot \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_t \end{bmatrix} \text{ dans l'espace fréquentiel.} \quad (2.10)$$

On retrouve bien l'expression classique du flot optique monocouche sur l'image I en écrivant $\mathbf{a}(\mathbf{w}) \cdot I(\mathbf{p}) = 0$ dans l'espace direct, et $\tilde{\mathbf{a}}(\mathbf{w}) \cdot I(\omega) = 0$ dans l'espace des fréquences. Sous l'hypothèse de petits mouvements, la résolution de ces équations permet d'estimer les vitesses sur des images sans transparence.

Si nous considérons maintenant une image composée de n couches I_i de vitesses \mathbf{w}_i :

$$I(\mathbf{p}) = \sum_{i=1}^n I_i(\mathbf{p}), \quad (2.11)$$

$$I(\omega) = \sum_{i=1}^n I_i(\omega). \quad (2.12)$$

En appliquant successivement les opérateurs de flot optique $\mathbf{a}(\mathbf{w}_1)$, $\mathbf{a}(\mathbf{w}_2)$... $\mathbf{a}(\mathbf{w}_n)$ à ces équations, nous établissons les égalités suivantes :

$$\mathbf{a}(\mathbf{w}_1)\mathbf{a}(\mathbf{w}_2)\dots\mathbf{a}(\mathbf{w}_n)I(\mathbf{p}) = \mathbf{a}(\mathbf{w}_1)\mathbf{a}(\mathbf{w}_2)\dots\mathbf{a}(\mathbf{w}_n)\sum_{i=1}^n I_i(\mathbf{p}) \quad (2.13)$$

$$= \sum_{i=1}^n \mathbf{a}(\mathbf{w}_1)\mathbf{a}(\mathbf{w}_2)\dots\mathbf{a}(\mathbf{w}_n) I_i(\mathbf{p}) \quad (2.14)$$

(par linéarité de l'opérateur de flot optique)

$$= \sum_{i=1}^n \left(\prod_{j=1}^n \mathbf{a}(\mathbf{w}_j) \right) I_i(\mathbf{p}) \quad (2.15)$$

(en introduisant une nouvelle notation)

$$= \sum_{i=1}^n \left(\prod_{j=1, j \neq i}^n \mathbf{a}(\mathbf{w}_j) \right) \mathbf{a}(\mathbf{w}_i) I_i(\mathbf{p}) \quad (2.16)$$

(par commutativité de l'opérateur flot optique)

$$= \sum_{i=1}^n \left(\prod_{j=1, j \neq i}^n \mathbf{a}(\mathbf{w}_j) \right) 0 \quad (2.17)$$

(par conservation de l'intensité de la couche i)

$$= 0 \text{ (par linéarité de l'opérateur de flot optique)}. \quad (2.18)$$

De façon similaire, on a :

$$\tilde{\mathbf{a}}(\mathbf{w}_1)\tilde{\mathbf{a}}(\mathbf{w}_2)\dots\tilde{\mathbf{a}}(\mathbf{w}_n)\tilde{I}(\omega) = 0. \quad (2.19)$$

Les équations (2.18) et (2.19) sont des équivalents à l'équation contrainte de mouvement classique, mais adaptés aux images transparentes. Des approches d'estimation de mouvements spécifiques aux mouvements transparents peuvent être construites en cherchant à résoudre l'une de ces deux équations sur les images considérées. Attardons-nous sur les hypothèses faites par les auteurs. Comme pour les méthodes de flot optique classiques, il faut que les mouvements soient petits, ou alors imaginer une approche multirésolution, pour que les équations (2.18) et (2.19) restent valides.

Surtout, pour chaque couche supplémentaire ajoutée au modèle de formation d'image transparente, il faut appliquer l'opérateur de flot optique une fois de plus. Ceci implique de manipuler des dérivées d'ordre supérieur. Une première conséquence est une plus forte dépendance au bruit entâchant les images. Mais il faut aussi calculer des dérivées temporelles d'ordre $n + 1$ s'il y a n couches, ce qui suppose que **le mouvement soit**

constant entre $n + 1$ instants successifs. C'est là bien sûr une hypothèse très forte sur laquelle nous reviendrons plus loin.

Barth *et al.* [Barth 03] et Yue *et al.* [Yu 03] ont étudié comment l'opérateur $\tilde{\mathbf{a}}(\mathbf{w}_1)\tilde{\mathbf{a}}(\mathbf{w}_2)\dots\tilde{\mathbf{a}}(\mathbf{w}_n)$ se comporte sur des images comportant des occultations (et non pas de la transparence). Ils montrent qu'un terme parasite de distorsion apparaît aux basses fréquences et empêche de s'appuyer sur (2.19).

2.1.2.2 Méthode de résolution

Shizawa et Mase ont donc été les premiers à formaliser complètement le problème de la transparence. Leur approche théorique a inspiré l'essentiel des travaux qui ont porté sur le problème.

Ils ont en outre développé un algorithme d'estimation des vitesses sur images transparentes dans leurs trois articles [Shizawa 90, Shizawa 91b, Shizawa 91a]. Pour résoudre (2.18) ou (2.19), on peut construire une énergie à minimiser sur l'image. En faisant intervenir des variables intermédiaires (les *mixed motion parameters*), l'expression devient quadratique (alors que la dépendance aux vitesses était de degré 4). Il s'agit alors d'estimer d'abord ces *mixed motion parameters*, avant de reconstruire les vitesses.

Les auteurs montrent dans leurs articles qu'il est équivalent de minimiser cette énergie et de trouver le vecteur propre d'une matrice particulière \mathbf{J}_n (dont ils donnent l'expression analytique), associé à la plus petite valeur propre. Une discussion sur la proximité à zéro de cette valeur propre par rapport aux autres donne d'ailleurs un critère sur le nombre de couches transparentes présentes dans l'image.

Leur article le mieux détaillé sur le sujet est [Shizawa 91a]. Il développe cet algorithme dans le cas de deux couches, mais toujours d'un point de vue théorique sans évoquer tous les aspects relatifs à son implémentation. Il fait en particulier intervenir 376 filtres de Gabor ! Il a été appliqué seulement sur des images très simples de bruits texturés animés de (petits) mouvements particulièrement réguliers.

Mota et Stuke proposent une méthode de résolution plus détaillée dans [Mota 01]. Comme Shizawa et Mase, ils veulent déterminer les *mixed motion parameters* par une procédure de diagonalisation : il s'agit du vecteur propre associé à la plus petite valeur propre d'une matrice \mathbf{J}_n . Celle-ci s'exprime de manière analytique en fonction des dérivées premières et secondes de l'image. Ils vont cependant plus loin dans la résolution de ce problème en montrant que le vecteur propre recherché s'exprime en fonction des mineurs² de \mathbf{J}_n .

Cette approche a l'avantage, comme celle de Shizawa et Mase, de permettre d'exploiter l'ordre de grandeur des différentes valeurs propres pour décider de la significativité de telle ou telle couche. Cependant, même si l'utilisation des mineurs allège les calculs, cette méthode reste théorique et son efficacité n'a été démontrée que sur des images très artificielles de bruits texturés. Enfin, l'absence de lissage explicite laisse augurer d'une très forte sensibilité au bruit temporel.

²Le mineur M_{ij} est le déterminant de \mathbf{J}_n privé de sa i -ème ligne et de sa j -ème colonne.

Globalement, si ces travaux ont été pionniers sur le problème de la mesure du mouvement avec transparence, et ont ouvert des voies très fructueuses, leurs propositions d'estimateurs sont moins convaincantes en pratique. D'autres travaux ont trouvé un équivalent non pas simplement à l'équation contrainte du mouvement, mais au principe même de la conservation d'intensité, ce qui permet de ne plus se limiter aux petits mouvements. La prochaine section décrit les approches de ce type proposées dans le cadre fréquentiel, et la suivante celles qui s'inscrivent dans l'espace direct de la séquence d'images.

2.2 Estimation de mouvements transparents dans l'espace de Fourier par généralisation de la conservation de l'intensité

2.2.1 Principe

Pingault, Pellerin et Shimahara [Pingault 02c, Pingault 04, Pingault 03a] proposent une relecture de la conservation de l'intensité dans le cadre fréquentiel. Il n'est alors plus nécessaire de se limiter aux petits mouvements.

Dans le cadre monocouche, la conservation de l'intensité s'écrit :

$$I(\mathbf{p}, t) = I(\mathbf{p} + \mathbf{w}, t - 1) \quad (2.20)$$

$$= I(x + u, y + v, t - 1) \quad (2.21)$$

avec $\mathbf{p} = (x, y)$ et $\mathbf{w} = (u, v)$, déplacement de \mathbf{p} de t vers $t - 1$. En supposant le mouvement constant depuis l'origine temporelle,

$$I(\mathbf{p}, t) = I(x + u.t, y + v.t, 0) \quad (2.22)$$

$$= I^0(x + u.t, y + v.t) \text{ en introduisant une nouvelle notation.} \quad (2.23)$$

Ce qui s'écrit, dans l'espace de Fourier :

$$I(\omega_x, \omega_y, \omega_t) = I^0(\omega_x, \omega_y) \delta(-u.\omega_x - v.\omega_y + \omega_t). \quad (2.24)$$

avec $\delta(\cdot)$ la fonction de Dirac. Autrement dit, la transformée de Fourier 3D de la séquence d'images I est nulle partout sauf dans le plan défini par :

$$u.\omega_x + v.\omega_y - \omega_t = 0. \quad (2.25)$$

Dans le cas de la transparence maintenant, chaque couche vérifiant (2.24), nous avons :

$$I(\omega_x, \omega_y, \omega_t) = \sum_{i=1}^n I_i(\omega_x, \omega_y, \omega_t) \quad (2.26)$$

$$= \sum_{i=1}^n I_i^0(\omega_x, \omega_y) \delta(-u.\omega_x - v.\omega_y + \omega_t). \quad (2.27)$$

La transformée de Fourier 3D de la séquence d'images transparentes est donc composée de plans, représentant chacun le spectre de Fourier 2D d'une couche i , qui a pour équation :

$$u_i.\omega_x + v_i.\omega_y - \omega_t = 0. \quad (2.28)$$

Si l'on sait estimer l'orientation de ces différents plans, on peut donc trouver les vitesses des couches correspondantes. Les approches proposées pour ce faire sont détaillées dans la sous-section suivante.

L'hypothèse la plus forte faite ici est que les mouvements doivent être non seulement translationnels spatialement (ou dans un voisinage spatial dans le cas de transformées de Fourier fenêtrées), mais surtout **constants sur plusieurs dates successives (en pratique au moins une dizaine)** pour permettre une extraction fiable des plans. Nous y reviendrons, car c'est la limitation essentielle de ce type d'approche.

2.2.2 Méthodes de résolution

Filtres orientés. L'approche la plus directe pour estimer les orientations de ces différents plans consiste à étudier la réponse de filtres orientés dans le spectre $3D$ de la séquence transparente. C'est ainsi que procèdent Yu *et al.* [Yu 03].

Ils balayent systématiquement l'espace des orientations angulaires (θ, ϕ) avec des filtres coniques, ce qui leur permet de calculer ce qu'ils appellent la signature de l'espace transformé. Ils gardent alors les orientations ayant donné lieu aux réponses les plus marquées.

Ils complètent leur algorithme par un test de fiabilité vérifiant si les orientations extraites sont véritablement significatives. Outre l'hypothèse de mouvement constant sur une longue fenêtre temporelle, on peut également opposer à ce type de méthode le temps de calcul important que suppose une telle recherche exhaustive.

Milanfar [Milanfar 96, Milanfar 99] propose pour réduire les coûts de calcul de ne pas explorer l'espace de Fourier $3D$ entier, mais de se limiter à un travail dans deux sous-espaces particuliers : (ω_x, ω_y) et (ω_x, ω_t) par exemple. Les intersections des plans correspondant aux couches avec ces deux sous-espaces suffisent en effet à définir parfaitement l'orientation des plans, et par suite les vitesses des couches.

Il n'est de plus pas nécessaire de calculer la transformée de Fourier $3D$ pour connaître ses valeurs dans les deux coupes (ω_x, ω_y) et (ω_x, ω_t) . La théorie des projections précise en effet que la transformée de Fourier de la projection de l'image suivant l'axe y par exemple, n'est autre que la coupe du spectre $3D$ de la séquence dans le plan (ω_x, ω_t) .

Dans un premier article [Milanfar 96], les droites sont détectées dans chacun des deux plans de coupe par une méthode dite SLIDE, plus rapide que la transformée de Hough. Il reste alors un problème d'appariement entre les droites détectées dans un plan et dans l'autre, que l'auteur règle en explorant tous les couples possibles.

Dans [Milanfar 99], l'auteur revient à la transformée de Hough pour détecter les droites, et il considère trois plans de coupe pour régler le problème d'appariement. Ce second algorithme a de meilleures performances que le précédent en termes de précision.

Spinéi et Pellerin [Spinei 01] se basent aussi sur une recherche d'orientation par utilisation de filtres orientés, mais ils se focalisent sur quelques configurations de transparence bien particulières, ce qui a pour conséquence de réduire drastiquement le volume de calcul à effectuer. Ils montrent que lorsque l'une des deux couches est fixe, ou que les deux couches ont un mouvement opposé, on peut se contenter d'analyser la réponse de trois filtres (les triades de Gabor) pour estimer les vitesses de ces couches.

Détection de plans dans l'espace de Fourier 3D. Une alternative est d'essayer de reconnaître directement les plans présents dans la transformée de Fourier 3D de la séquence. Elle a été pour la première fois esquissée par Yu *et al.* [Yu 99]. Les auteurs emploient un algorithme EM pour déterminer l'orientation des plans en présence. Cependant, la majeure partie de leur article étant consacrée à la rectification des distorsions dues aux occultations, l'estimation des mouvements transparents est peu détaillée et les exemples donnés sont peu explicites.

Pingault opte pour le même type d'approche dans la première partie de sa thèse [Pingault 03a]. Il y développe deux méthodes d'estimation permettant de calculer l'orientation des plans : la première fait intervenir une minimisation EM, l'autre une minimisation SEM.

Il propose deux prétraitements avant tout. Puisque les plans passent tous par l'origine, ils sont difficiles à distinguer aux basses fréquences. Il applique donc un passe-haut sur le spectre de la séquence pour réhausser les hautes fréquences, les plus intéressantes mais les plus rares dans les séquences naturelles. En outre, les spectres étant bruités, il en extrait seulement les pics les plus significatifs, pour finalement ne plus travailler que sur 300 points environ. Cela a en outre l'avantage de réduire la masse de données à manipuler. Le problème est alors de déterminer le nombre et l'orientation des plans qui donnent lieu à ce nuage de points.

L'estimation par méthode EM (*Expectation Maximisation*) est détaillée dans sa thèse [Pingault 03a]. Pingault modélise le nuage des points extraits comme un mélange de n plans, dont chacun définit une densité de probabilité $f_k(\mathbf{x}|\theta_k)$ (θ_k est le vecteur de paramètres qui définit l'orientation du plan et son épaisseur) :

$$f(\mathbf{x}|\theta) = \sum_{k=1}^p \alpha_k f_k(\mathbf{x}|\theta_k) \text{ avec } \sum_{k=1}^p \alpha_k = 1. \quad (2.29)$$

Il choisit pour $f_k(x|\theta_k)$ des gaussiennes 3D. Les deux plus grandes valeurs propres de leur matrice de covariance définiront l'orientation du plan et la troisième son épaisseur. L'estimation se fait itérativement :

- On calcule la probabilité pour chaque point \mathbf{x}_i d'appartenir à chacune des gaussiennes 3D.
- On estime alors les poids α_k , puis les covariances des gaussiennes.

Tout ceci est exprimé dans un formalisme EM. Le choix du nombre de plans se fait au départ en surestimant le nombre de couches attendues. Les auteurs éliminent *a posteriori* celles qui sont peu significatives. Seulement, la modélisation par des gaussiennes 3D pose des problèmes de stabilité lorsque les plans sont bien définis et que la plus petite valeur propre est alors trop proche de 0. D'autre part, l'algorithme a tendance à fusionner les plans trop proches.

C'est pourquoi les auteurs proposent une variante plus stable de cette méthode [Pingault 02c, Pingault 03a, Pingault 04]. Les plans sont alors estimés par l'algorithme SEM (*Stochastic Expectation Maximisation*). Au lieu de calculer pendant l'étape E la probabilité $p_{k,i}$ de chaque point \mathbf{x}_i d'appartenir à une gaussienne k , on attribue \mathbf{x}_i à un plan (et un seul), en échantillonnant une loi multinômiale suivant les $p_{k,i}$. En outre, le modèle

probabiliste est modifié pour intégrer un *a priori*. Les auteurs robustifient leur approche en intégrant à leur modèle le fait que chaque plan donne lieu à des fantômes qui lui sont parallèles (que Pingault appelle *plans de repliement*), et en introduisant une classe de rejet. Les expériences montrent que cette nouvelle approche est performante, et robuste à certains cas particuliers (transparence triple par endroits, variation de l'intensité). Par contre, l'hypothèse très forte de constance du mouvement sur un grand nombre d'images (15 ici) reste présente.

Un autre formalisme Citons brièvement les travaux inclassables de Darrell et Simoncelli ([Darrell 91], [Simoncelli 92] et [Darrell 93]). Pour intégrer le fait que les vitesses sont souvent déterminées de manière ambiguë dans un voisinage local, les auteurs proposent une alternative aux champs de vitesse : des profils de vitesses, qui sont assimilables à des probabilités de vitesse sur le voisinage. Ils proposent différentes méthodes pour les calculer : à l'aide de filtres *donuts* (nommés ainsi en raison de leur forme) ou *nulling filters* travaillant dans l'espace de Fourier. Ce modèle probabiliste a l'avantage de permettre d'introduire des *a priori* dans un formalisme bayésien, mais doit être seuillé pour fournir un champ de vitesse exploitable en fin d'algorithme. Comme les vitesses sont déterminées par des filtres dans l'espace de Fourier, les mouvements transparents y sont également détectés. L'approche suit cependant un formalisme très particulier. Il faut également remarquer que la recherche des vitesses se fait de manière systématique dans l'espace de Fourier.

Conclusion partielle Finalement, les approches fréquentielles sont intéressantes dans la mesure où elles modélisent exactement le phénomène de transparence, et permettent donc des estimations fiables. De plus, comme nous le verrons dans la fin de ce chapitre, elles permettent de séparer les couches (dans des conditions idéales). En effet, en séparant les différents plans dans l'espace de Fourier $3D$, on peut retrouver la transformée de Fourier des couches en question (comme le montre l'équation (2.27)), et donc de reconstruire chaque couche.

Il n'est cela dit pas nécessaire de rentrer dans le détail des différentes méthodes évoquées pour réaliser qu'elles ne sont pas applicables pour notre problème. Comme déjà évoqué, elles supposent des mouvements des couches constants sur plusieurs images (au moins une dizaine en pratique), ce qui n'est pas acceptable dans notre cas. (À 30 images par seconde, cela revient à supposer une constance des mouvements sur 0.5 s environ, ce qui représente un demi-cycle cardiaque.)

2.3 Estimation de mouvements transparents dans l'espace direct

2.3.1 Principe

Etude de l'équation. Il est également possible d'établir directement un équivalent à l'équation de conservation de l'intensité. Considérons une séquence d'images transparentes I , superposition de deux couches I_1 et I_2 . En tout pixel \mathbf{p} à la date t , on peut écrire :

$$I(\mathbf{p}, t) = I_1(\mathbf{p}, t) + I_2(\mathbf{p}, t). \quad (2.30)$$

Soient $\mathbf{w}_1(\mathbf{p}, t)$ et $\mathbf{w}_2(\mathbf{p}, t)$ les vitesses des deux couches au pixel \mathbf{p} et à la date t . La conservation de l'intensité entre les dates t et $t + 1$ *au niveau des couches* s'écrit en tout point :

$$I_1(\mathbf{p}, t + 1) = I_1(\mathbf{p} + \mathbf{w}_1(\mathbf{p}, t), t), \quad (2.31)$$

$$I_2(\mathbf{p}, t + 1) = I_2(\mathbf{p} + \mathbf{w}_2(\mathbf{p}, t), t). \quad (2.32)$$

Formons maintenant la quantité suivante :

$$\begin{aligned} & I(\mathbf{p}, t + 1) - I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}, t), t) - I(\mathbf{p} + \mathbf{w}_2(\mathbf{p}, t), t) \\ & + I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}, t - 1) + \mathbf{w}_2(\mathbf{p}, t - 1), t - 1) \\ & = \left(I_1(\mathbf{p}, t + 1) + I_2(\mathbf{p}, t + 1) \right) - \left(I_1(\mathbf{p} + \mathbf{w}_1(\mathbf{p}, t), t) + I_2(\mathbf{p} + \mathbf{w}_1(\mathbf{p}, t), t) \right) \\ & \quad - \left(I_1(\mathbf{p} + \mathbf{w}_2(\mathbf{p}, t), t) + I_2(\mathbf{p} + \mathbf{w}_2(\mathbf{p}, t), t) \right) + \left(I_1(\mathbf{p} + \mathbf{w}_1(\mathbf{p}, t - 1) + \mathbf{w}_2(\mathbf{p}, t - 1), t - 1) \right) \\ & \quad + I_2(\mathbf{p} + \mathbf{w}_1(\mathbf{p}, t - 1) + \mathbf{w}_2(\mathbf{p}, t - 1), t - 1) \end{aligned} \quad (2.33)$$

en décomposant l'image en ses deux couches sous-jacentes.

Il est important de garder à ce stade des notations complètes, même si elles sont un peu lourdes, pour bien comprendre quelles approximations sont faites. En réorganisant l'expression précédente :

$$\begin{aligned} & I(\mathbf{p}, t + 1) - I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}, t), t) - I(\mathbf{p} + \mathbf{w}_2(\mathbf{p}, t), t) \\ & + I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}, t - 1) + \mathbf{w}_2(\mathbf{p}, t - 1), t - 1) \\ & = \left(I_1(\mathbf{p}, t + 1) - I_1(\mathbf{p} + \mathbf{w}_1(\mathbf{p}, t), t) \right) \\ & \quad + \left(I_2(\mathbf{p}, t + 1) - I_2(\mathbf{p} + \mathbf{w}_2(\mathbf{p}, t), t) \right) \\ & \quad - \left(I_2(\mathbf{p} + \mathbf{w}_1(\mathbf{p}, t), t) - I_2((\mathbf{p} + \mathbf{w}_1(\mathbf{p}, t - 1)) + \mathbf{w}_2(\mathbf{p}, t - 1), t - 1) \right) \\ & \quad - \left(I_1(\mathbf{p} + \mathbf{w}_2(\mathbf{p}, t), t) - I_1((\mathbf{p} + \mathbf{w}_2(\mathbf{p}, t - 1)) + \mathbf{w}_1(\mathbf{p}, t - 1), t - 1) \right) \end{aligned} \quad (2.34)$$

$$\begin{aligned} & = - \left(I_2(\mathbf{p} + \mathbf{w}_1(\mathbf{p}, t), t) - I_2((\mathbf{p} + \mathbf{w}_1(\mathbf{p}, t - 1)) + \mathbf{w}_2(\mathbf{p}, t - 1), t - 1) \right) \\ & \quad - \left(I_1(\mathbf{p} + \mathbf{w}_2(\mathbf{p}, t), t) - I_1((\mathbf{p} + \mathbf{w}_2(\mathbf{p}, t - 1)) + \mathbf{w}_1(\mathbf{p}, t - 1), t - 1) \right). \end{aligned} \quad (2.35)$$

Nous avons éliminé les deux premiers termes en reconnaissant la conservation de l'intensité *au niveau des couches*. Pour faire disparaître les deux derniers termes, il faut faire deux hypothèses.

La première suppose **les mouvements sont constants entre $t - 1$ et $t + 1$ pour les deux couches**. Même si cette contrainte est beaucoup plus limitée dans le temps que celle nécessaire à une estimation fréquentielle fiable, cela reste une hypothèse importante. Il est possible qu'elle ne soit pas vérifiée dans certaines applications. En ce qui nous concerne, elle sera problématique à certaines dates du cycle cardiaque. Nous y reviendrons.

L'équation (2.35) devient sous cette première hypothèse :

$$\begin{aligned}
 & I(\mathbf{p}, t + 1) - I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}, t), t) - I(\mathbf{p} + \mathbf{w}_2(\mathbf{p}, t), t) \\
 & + I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}, t - 1) + \mathbf{w}_2(\mathbf{p}, t - 1), t - 1) \\
 & = - \left(I_2(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t) - I_2((\mathbf{p} + \mathbf{w}_1(\mathbf{p})) + \mathbf{w}_2(\mathbf{p}), t - 1) \right) \\
 & \quad - \left(I_1(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t) - I_1((\mathbf{p} + \mathbf{w}_2(\mathbf{p})) + \mathbf{w}_1(\mathbf{p}), t - 1) \right). \tag{2.36}
 \end{aligned}$$

L'indexation des vitesses par les dates a disparu, mais nous devons faire une seconde hypothèse pour pouvoir appliquer la conservation de l'intensité à chacun des deux termes. Concentrons-nous sur le premier. Nous voudrions appliquer la conservation de l'intensité à la couche I_2 , entre les dates $t - 1$ et t , *au pixel* $\mathbf{p} + \mathbf{w}_1(\mathbf{p})$. Pour ce faire, il nous faut supposer :

$$\mathbf{w}_2(\mathbf{p}) = \mathbf{w}_2(\mathbf{p} + \mathbf{w}_1(\mathbf{p})). \tag{2.37}$$

Et de la même façon sur la couche I_1 :

$$\mathbf{w}_1(\mathbf{p}) = \mathbf{w}_1(\mathbf{p} + \mathbf{w}_2(\mathbf{p})). \tag{2.38}$$

Autrement dit, les mouvements des deux couches doivent être non seulement constants dans le temps entre $t - 1$ et $t + 1$, mais **ils doivent aussi être constants spatialement à l'échelle du déplacement de l'autre couche.**

Si la première hypothèse, la plus lourde en pratique, a bien été notée par tous les auteurs s'étant intéressés à la question, la seconde n'a jamais été soulignée. Elle a pourtant pour conséquence qu'en théorie, *seule la classe des translations vérifie exactement la généralisation de la conservation de l'intensité pour la transparence.*

En fait, nous verrons plus loin que les algorithmes d'estimation de mouvements transparents ne souffrent pas de la violation de cette hypothèse. Ceci s'explique par le fait que les mouvements anatomiques sont faiblement divergents, et que nos couches sont peu contrastées.

Ceci étant précisé, l'équation recherchée s'écrit comme suit :

$$\begin{aligned}
 & I(\mathbf{p}, t + 1) - I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t) \\
 & - I(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t) + I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}) + \mathbf{w}_2(\mathbf{p}), t - 1) = 0. \tag{2.39}
 \end{aligned}$$

Il s'agit bien d'une généralisation de la conservation de l'intensité dans le sens où cette équation ne fait aucune hypothèse sur l'amplitude des déplacements et ne met en jeu que les images de la séquence (sur un intervalle de trois instants).

Etablissement de l'équation (2.39) Cela dit, l'équation (2.39) n'a pas le sens physique clair que porte l'hypothèse usuelle de conservation de l'intensité. Cette dernière traduit en effet le fait qu'un pixel conserve son intensité le long de sa trajectoire dans le plan images. L'équation (2.39) n'a malgré tout pas été établie de manière heuristique. Pour y aboutir, Pingault et Pellerin [Pingault 02b, Pingault 03a] ont écrit les équations de superposition des couches (2.30), et de conservation de l'intensité des couches (2.31) et (2.32) dans l'espace de Fourier, en représentant le déplacement des couches par des opérateurs de

décalage $\Delta_1(\omega_x, \omega_y)$ et $\Delta_2(\omega_x, \omega_y)$ et en notant $I^t(\omega_x, \omega_y)$ la transformée de Fourier 2D de la séquence prise à la date t :

$$I^t(\omega_x, \omega_y) = I_1^{t-1}(\omega_x, \omega_y) \cdot \Delta_1(\omega_x, \omega_y) + I_2^{t-1}(\omega_x, \omega_y) \cdot \Delta_2(\omega_x, \omega_y). \quad (2.40)$$

En écrivant cette équation à deux dates successives, il est possible de faire disparaître les termes $I_1^{t-1}(\omega_x, \omega_y)$ et $I_2^{t-1}(\omega_x, \omega_y)$ des équations :

$$\begin{aligned} I^{t+1}(\omega_x, \omega_y) &= I^t(\omega_x, \omega_y) \cdot \Delta_1^{-1}(\omega_x, \omega_y) + I^t(\omega_x, \omega_y) \cdot \Delta_2^{-1}(\omega_x, \omega_y) \\ &- I^{t-1}(\omega_x, \omega_y) \cdot \Delta_1^{-1}(\omega_x, \omega_y) \cdot \Delta_2^{-1}(\omega_x, \omega_y). \end{aligned} \quad (2.41)$$

On retrouve bien (2.39) par transformée de Fourier inverse. Pingault souligne d'ailleurs dans sa thèse [Pingault 03a] qu'on peut de la même façon établir la conservation de l'intensité à des transparences à n couches, à condition de supposer le mouvement constant sur $n+1$ dates successives :

$$\sum_{i=t+1-n}^{t+1} (-1)^{t+1-i} \sum_{\{j_1, \dots, j_{t+1-i}\} \in \sigma_{t+1-i}([1, n])} I(\mathbf{p} + \mathbf{w}_{j_1} + \dots + \mathbf{w}_{j_{t+1-i}}, i) = 0 \quad (2.42)$$

où $\sigma_j([1, n])$ les permutations possibles de j éléments de $[1, n]$. Cette démonstration suppose que l'on peut commuter les opérateurs $\Delta_1(\omega_x, \omega_y)$ et $\Delta_2(\omega_x, \omega_y)$, ce qui n'est vrai en toute rigueur que quand les vitesses sont translationnelles.

Linéarisations. L'expression (2.39) est difficile à manipuler telle quelle puisqu'elle est non linéaire. C'est pourquoi une étape de linéarisation est introduite dans la phase de minimisation. Nous mènerons les calculs qui nous intéresseront dans le chapitre 3. Ils sont légèrement différents en ce que nous n'effectuons la linéarisation que dans l'espace (et pas dans le temps).

Faisons néanmoins deux remarques dès à présent :

- Il est souvent nécessaire de linéariser l'expression (2.39) au second ordre. La linéarisation au premier ordre présente en effet une singularité en 0, et il faut donc monter d'un ordre pour avoir une expression non dégénérée.
- Une linéarisation de (2.39) au second ordre en 0 permet de retrouver l'expression du flot optique généralisé (2.18). En effet, en notant ∂_t le dérivée temporelle, ∇I le gradient de I et $\mathbf{H}(I)$ son hessien :

$$\begin{aligned} &I(\mathbf{p}, t+1) - I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t) - I(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t) + I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}) + \mathbf{w}_2(\mathbf{p}), t-1) \\ &= \left(I(\mathbf{p}, t) + \partial_t I(\mathbf{p}, t) + \frac{1}{2} \cdot \partial_{t^2} I(\mathbf{p}, t) \right) \\ &\quad - \left(I(\mathbf{p}, t) + \mathbf{w}_1(\mathbf{p}) \nabla I(\mathbf{p}, t) + \frac{1}{2} \cdot \mathbf{w}_1(\mathbf{p})^T \mathbf{H}(I)(\mathbf{p}, t) \mathbf{w}_1(\mathbf{p}) \right) \\ &\quad - \left(I(\mathbf{p}, t) + \mathbf{w}_2(\mathbf{p}) \nabla I(\mathbf{p}, t) + \frac{1}{2} \cdot \mathbf{w}_2(\mathbf{p})^T \mathbf{H}(I)(\mathbf{p}, t) \mathbf{w}_2(\mathbf{p}) \right) \\ &\quad + \left(I(\mathbf{p}, t) + (\mathbf{w}_1(\mathbf{p}) + \mathbf{w}_2(\mathbf{p})) \nabla I(\mathbf{p}, t) \right) \\ &\quad + \frac{1}{2} (\mathbf{w}_1(\mathbf{p}) + \mathbf{w}_2(\mathbf{p}))^T \mathbf{H}(I)(\mathbf{p}, t) (\mathbf{w}_1(\mathbf{p}) + \mathbf{w}_2(\mathbf{p})) \\ &\quad - (\mathbf{w}_1(\mathbf{p}) + \mathbf{w}_2(\mathbf{p})) \nabla \partial_t I(\mathbf{p}, t) - \partial_t I(\mathbf{p}, t) + \frac{1}{2} \cdot \partial_{t^2} I(\mathbf{p}, t). \end{aligned} \quad (2.43)$$

Les trois premières lignes correspondent à la linéarisation des trois premiers termes respectivement, et les trois dernières au dernier terme (qu'il faut linéariser à la fois spatialement et temporellement).

Pour alléger les notations, on ne mentionnera plus le pixel \mathbf{p} ni la date t . En réorganisant (2.43) :

$$\begin{aligned} & I(\mathbf{p}, t + 1) - I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t) - I(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t) + I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}) + \mathbf{w}_2(\mathbf{p}), t - 1) \\ &= \partial_{t^2} I + \frac{1}{2} \cdot \mathbf{w}_2^T \mathbf{H}(I) \mathbf{w}_1 + \frac{1}{2} \cdot \mathbf{w}_1^T \mathbf{H}(I) \mathbf{w}_2 - (\mathbf{w}_1 + \mathbf{w}_2) \nabla \partial_t I. \end{aligned} \quad (2.44)$$

Notons que les termes du premier ordre ont bien disparu. En faisant apparaître les composantes en x et en y :

$$\begin{aligned} & I(\mathbf{p}, t + 1) - I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t) - I(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t) + I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}) + \mathbf{w}_2(\mathbf{p}), t - 1) \\ &= \partial_{t^2} I + 1/2 \cdot u_2 \cdot u_1 \cdot \partial_{x^2} I + 1/2 \cdot v_2 \cdot v_1 \cdot \partial_{y^2} I + 1/2 \cdot u_2 \cdot v_1 \cdot \partial_{xy} I \\ &\quad + 1/2 \cdot v_2 \cdot u_1 \cdot \partial_{xy} I + 1/2 \cdot u_1 \cdot u_2 \cdot \partial_{x^2} I + 1/2 \cdot v_1 \cdot v_2 \cdot \partial_{y^2} I \\ &\quad + 1/2 \cdot u_1 \cdot v_2 \cdot \partial_{xy} I + 1/2 \cdot v_1 \cdot u_2 \cdot \partial_{xy} I \\ &\quad - (u_1 + u_2) \cdot \partial_{xt} I - (v_1 + v_2) \cdot \partial_{yt} I \end{aligned} \quad (2.45)$$

$$\begin{aligned} &= \partial_{t^2} I + u_1 \cdot u_2 \cdot \partial_{x^2} I + v_1 \cdot v_2 \cdot \partial_{xy} I + u_1 \cdot v_2 \cdot \partial_{xy} I \\ &\quad + v_1 \cdot u_2 \cdot \partial_{y^2} I - (u_1 + u_2) \cdot \partial_{xt} I - (v_1 + v_2) \cdot \partial_{yt} I \end{aligned} \quad (2.46)$$

$$= (u_1 \partial_x + v_1 \partial_y - \partial_t) (u_2 \partial_x + v_2 \partial_y - \partial_t) I \quad (2.47)$$

$$= 0. \quad (2.48)$$

Nous reconnaissons bien en (2.47) l'équation du flot optique généralisé à la transparence (2.18). La généralisation de la conservation de l'intensité mène dans le cas particulier d'une linéarisation spatio-temporelle autour du point $(\mathbf{p}, t + 1)$ à l'équation établie par Shizawa et Mase.

C'est pourquoi ces deux approches peuvent se recouper fortement. Nous avons cependant pour des raisons didactiques et historiques décidé de les séparer dans le plan de cette bibliographie. Nous avons pris le parti de réduire la présentation des méthodes appliquant le flot optique généralisé aux approches découlant strictement de l'approche de Shizawa et Mase.

2.3.2 Méthodes de résolution

De nombreuses méthodes, inspirées des approches suivies en estimation de mouvement, ont été proposées pour résoudre l'équation (2.39) sur toute l'image. Elles cherchent à trouver le champ des vitesses $(\mathbf{w}_1(\mathbf{p}), \mathbf{w}_2(\mathbf{p}))$ tel qu'une énergie du type :

$$\begin{aligned} J(\mathbf{w}_1, \mathbf{w}_2) &= \sum_{p \in \mathfrak{S}} \left(I(\mathbf{p}, t + 1) - I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t) - I(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t) \right. \\ &\quad \left. + I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}) + \mathbf{w}_2(\mathbf{p}), t - 1) \right)^2. \end{aligned} \quad (2.49)$$

soit minimisée, où \mathfrak{S} représente les points de la grille image.

Nous pouvons distinguer deux grands types d'approches :

- Certains auteurs utilisent un modèle des champs de vitesses, spécifié par un jeu de paramètres qu'il s'agit d'estimer. Plus le modèle est contraignant, moins il sera généralisable à une large classe de mouvements transparents. Par contre, les mouvements conformes au modèle seront estimés rapidement, et avec une grande robustesse au bruit.
- On peut aussi chercher à calculer un champ dense de vitesses, ce qui permettra d'attaquer des problèmes de transparence très généraux. Cela pose néanmoins des problèmes de temps de calcul et de robustesse au bruit, et il faut définir avec soin un terme de régularisation à adjoindre à l'équation (2.49) pour apporter une contrainte appropriée à la solution attendue.

Nous présentons maintenant les contributions correspondant à ces deux types de méthodes, ainsi qu'une approche intermédiaire intéressante.

Modélisation paramétrique des champs de vitesses. Une première technique, proposée par Stuke *et al.* ([Stuke 03a, Stuke 04]), suppose le mouvement translationnel par blocs. Il s'agit donc de minimiser sur chaque bloc B l'énergie :

$$\min_{\mathbf{w}_1, \mathbf{w}_2} \sum_{p \in B} \left(I(\mathbf{p}, t+1) - I(\mathbf{p} + \mathbf{w}_1, t) - I(\mathbf{p} + \mathbf{w}_2, t) + I(\mathbf{p} + \mathbf{w}_1 + \mathbf{w}_2, t-1) \right)^2 \quad (2.50)$$

Le calcul de la solution se fait en testant toutes les valeurs possibles dans un certain intervalle. Cela permet de s'affranchir de problèmes de minimisation par technique de gradients (risquant d'aboutir à des minima locaux), et les calculs en question sont rapides³ puisqu'ils ne demandent aucune interpolation pour des valeurs entières de déplacement.

Les inconvénients viennent par contre du fait que

- si l'on choisit de blocs de taille substantielle, la contrainte de translation locale devient très forte ;
- si l'on choisit de petits blocs, on s'expose au bruit et à des problèmes d'ouverture.

Dans leurs articles, les auteurs proposent une mise en correspondance par blocs pour des transparences de n couches, et ils imaginent également une manière d'arbitrer entre transparence et occultation. Les seules images traitées sont très texturées, ce qui permet d'utiliser de petits blocs (5×5) sans risque de blocs contenant trop peu d'information exploitable.

Pingault et Pellerin ont fait eux aussi l'hypothèse de mouvement constant par blocs dans [Pingault 02b] et [Pingault 03a], mais ils ont proposé une méthode de résolution différente, qui ne nécessite pas de supposer les déplacements entiers.

Ils utilisent une méthode incrémentale classique de type Gauss-Newton, consistant à affiner successivement les estimés. Pour ce faire, ils linéarisent l'expression (2.50) autour de l'estimé courant des vitesses, ce qui leur permet d'estimer un correctif pour chacune des vitesses en minimisant l'expression obtenue.

Cette approche est classiquement utilisée lors d'estimation de mouvements monocouches. Deux différences majeures apparaissent quand on considère le problème de la transparence :

³Ils ne sont pas aussi rapides cependant que la mise en correspondance par bloc monocouche car il faut balayer un espace de dimension 4.

- Comme souligné plus haut, du fait d'une singularité en 0, on ne peut se contenter de linéariser (2.50) au premier ordre. Il faut pousser jusqu'au second ordre pour obtenir une expression bien définie quelle que soit la valeur de l'estimé précédent.
 - La fonction à minimiser devient alors de degré 4. Les auteurs substituent des inconnues supplémentaires aux termes de degré supérieur à 2, ce qui leur permet de remplacer une expression de degré 4 à 4 inconnues par une expression de degré 2 à 8 inconnues. Ces variables intermédiaires ne sont autres que les *mixed motion parameters* introduits par Shizawa et Mase.
- Cette nouvelle formulation du problème se résoud alors par une inversion matricielle. Nous verrons dans le chapitre suivant que nous avons préféré garder l'expression de degré 4 et la minimiser itérativement.

Ce principe est mis en oeuvre dans un cadre multirésolution et robuste. Le premier point permet de prendre en compte de larges déplacements, et le second de ne pas laisser des mouvements secondaires fausser l'estimation. La linéarisation au second ordre n'est utilisée qu'aux bas niveaux de la pyramide. En effet, une fois l'ambiguïté en 0 levée, il est possible de se contenter du premier ordre.

La principale limitation de cette méthode vient du fait qu'un voisinage de grande taille (21×21) est nécessaire pour avoir un système suffisamment complet d'équations permettant d'estimer correctement les 8 inconnues, ce qui est pénalisant dans le cas de mouvements divergents. En outre, l'introduction des *mixed motion parameters* découple artificiellement certaines variables et modifie donc légèrement la formulation du problème.

Pingault et Pellerin ont développé une deuxième méthode dans l'espace direct [Pingault 02a, Pingault 03a, Pingault 03b], qui se différencie essentiellement par le fait que le mouvement n'est plus modélisé par des translations locales indépendantes, mais est décomposé sur une base de B-splines, qui s'avère une représentation beaucoup plus flexible. Par contre, la robustesse au bruit et le temps de calcul seront moins favorables.

L'expression (2.49) devient ainsi une fonction des coefficients associés à la base de splines. Ils sont estimés par une minimisation sur une version linéarisée de (2.49). Notons d'ailleurs que lorsque la linéarisation est effectuée au second ordre, les auteurs ne réduisent plus le problème à une équation de degré 2, mais effectuent la minimisation directement sur le polynôme de degré 4 à l'aide de l'algorithme de Levenberg-Marquardt. Ce principe est à nouveau mis en oeuvre dans un cadre multirésolution (auxquelles les splines se plient de manière élégante) et robuste.

Finalement, il s'agit d'un modèle attractif en ce qu'il contraint les mouvements sans poser d'hypothèse forte sur leur forme. Nous verrons que nous ferons cependant d'autres choix, car la méthode développée doit être robuste à un niveau de bruit conséquent dans les images fluoroscopique, et nous pouvons contraindre plus fortement les modèles de mouvements pour les mouvements anatomiques.

Calcul de champs denses de vitesses. Nous considérons maintenant l'utilisation de (2.39) pour le calcul de champs de vitesses transparents denses, qui a surtout intéressé Ingo Stuke, Cicero Mota, Til Aach et leurs équipes. Eux aussi s'appuient sur une version de (2.49) linéarisée au second ordre.

Dans de premiers articles [Stuke 03c, Stuke 03b, Mota 05a, Mota 05b], les auteurs considèrent cette version linéarisée de (2.49) comme le terme d'attache aux données d'une

énergie à minimiser. Ils choisissent de simplifier le problème en faisant intervenir des variables intermédiaires (les *mixed motion parameters* c_i) conduisant à une expression de degré 2, et non pas de degré 4. L'idée consiste toujours à estimer ces variables intermédiaires, pour ensuite calculer les mouvements transparents. Cette dernière étape est résolue élégamment en exprimant les vitesses recherchées comme les racines d'un polynôme dont les coefficients dépendent des c_i .

Le terme d'attache aux données est accompagné d'un terme de lissage pénalisant les variations des c_i . Il s'agit de la somme au carré des dérivées des *mixed motion parameters*. Une formulation de ce type mène à un système d'Euler-Lagrange d'équations différentielles linéaires, semblables aux équations obtenues pour l'estimation de champs denses dans le cas mono-mouvement.

Si l'étude théorique effectuée par les auteurs est intéressante, les exemples donnés ne permettent pas de se faire une idée de la qualité de l'estimation. Il est en outre probablement délicat de régler de manière générale le poids relatif entre attache aux données et lissage. Enfin, les temps de calcul sont importants.

Dans [Stuke 04], les auteurs font intervenir un formalisme markovien. Il permet d'estimer des champs denses sur des images bruitées. En chaque site (ici les pixels) trois inconnues sont définies : les deux vecteurs de vitesse, et une variable binaire indiquant s'il y a un ou deux mouvements au pixel considéré. La vraisemblance conditionnelle des observations (trois images consécutives), à savoir le terme d'attache aux données, résulte de la conservation d'intensité (monocouche ou bicouche, selon les valeurs prises par la variable indicatrice binaire). *L'a priori* quant à lui vise à favoriser les situations où les sites voisins ont le même nombre de couches. Il cherche également à pénaliser des différences de vitesses importantes entre voisins. On peut également introduire de manière facultative un lissage temporel à partir des estimations obtenues aux dates précédentes.

La minimisation est effectuée soit par une variante de l'ICM, soit par un recuit simulé. Si les auteurs proposent une méthode originale, ils commentent peu son utilisation sur des exemples réels, et n'indiquent pas de temps de calcul (sans doute élevés), ni n'évoquent le choix des différents paramètres. Par contre, soulignons que le formalisme proposé gère par construction la segmentation des images en différentes zones, et reconnaît automatiquement les régions monocouches.

Une approche intermédiaire. Nous terminons le panorama des méthodes d'estimation de mouvements transparents basés sur la conservation d'intensité généralisée avec l'algorithme proposé par Toro *et al.* [Toro 00a, Toro 00b]. Il s'agit d'une méthode intermédiaire entre champ paramétré et champ dense, dans le sens où les auteurs commencent par une estimation dense des mouvements transparents, avant de mener une segmentation de ces champs en différentes zones où des modèles de mouvement affines sont introduits.

Pour le calcul du champ dense, la formulation de l'énergie à minimiser est classique : un terme d'attache aux données correspondant à (2.49) et un terme de lissage sur la différence entre deux vecteurs voisins. Des fonctions robustes sont préférées aux fonctions quadratiques pour chacun de ces termes.

La résolution du problème commence par une classique recherche des vecteurs de vitesse sous l'hypothèse de translations par blocs (sans terme de lissage). Les deux vitesses sont ensuite réestimées en chaque point de manière séquentielle et (*et non simultanément!*). La

dépendance vis à vis de la correction des premiers estimés est donc très forte.

En outre, les auteurs minimisent directement la fonction de coût non linéaire construite à l'aide d'un algorithme du *downhill simplex*. Cette méthode de minimisation itérative a l'avantage de bien fonctionner sur des fonctions complexes, mais elle est chère en temps de calcul, et ne garantit pas qu'un minimum pertinent soit trouvé.

La méthode d'estimation proposée prête donc à question. Par contre, l'étude du regroupement des vecteurs ainsi estimés en différents champs (affines ici) pour arriver à une segmentation des couches, est tout à fait intéressante. Un formalisme EM est exploité à cette fin :

- Pendant la phase M, on suppose la segmentation des champs connue, et on estime les paramètres des modèles de mouvement considérés. Les auteurs utilisent à nouveau le *downhill simplex* pour ce faire...
- Pendant la phase E, on détermine la carte de segmentation de sorte que chaque vecteur estimé dans la première phase de l'algorithme soit attribué au champ dont il est le plus proche.

Des estimateurs robustes sont introduits. Une classe de rejet est également créée : elle regroupe les pixels qui ne sont pas attribués à des régions segmentées.

Cette approche semble donc imparfaite, car les méthodes d'estimations employées n'apparaissent pas optimales. Les exemples décrits sont d'ailleurs peu convaincants. Il y a plusieurs paramètres à régler (un par fonction robuste, plus le poids relatif entre attache aux données et lissage). Le temps de calcul, comme la robustesse au bruit, n'est pas évoqué.

Nous avons malgré tout choisi de développer la présentation de ce travail dans notre état de l'art, car cette idée de procéder en deux étapes, en menant une segmentation au niveau de l'image dans la seconde, nous a beaucoup inspiré dans nos travaux sur la *bitransparence distribuée* (présentée dans le chapitre 5).

Conclusion sur l'approche dans l'espace direct. Comme les chapitres à venir le montreront, nous nous appuyerons exclusivement sur l'équation de la conservation de l'intensité généralisée (2.39) pour estimer les mouvements transparents dans les séquences d'images par rayons X. Elle procure en effet un équivalent à l'équation fondamentale qu'est la conservation de l'intensité dans le cas monocouche, ce qui permet d'adapter certaines des nombreuses méthodes développées dans ce cadre.

Les hypothèses permettant d'établir (2.39) sont beaucoup moins contraignantes que celles impliquées par les approches fréquentielles. Pour ce qui nous concerne, nous privilégierons les méthodes robustes au bruit important entachant les images fluoroscopiques, tout en évitant des coûts de calcul prohibitifs qui rendraient irréaliste toute implantation future en temps réel.

2.4 La problématique de la séparation des couches

Nous concluons cette bibliographie par une partie consacrée au problème connexe de la séparation des couches. Il s'agit d'une tâche très difficile à résoudre dans le cas général. C'est pourquoi, bien que les deux problématiques soient de toute évidence liées, la plupart des études menées en analyse de mouvements transparents les découplent.

Il est cependant intéressant d'examiner les approches proposées pour résoudre ce problème.

2.4.1 Formulation dans l'espace des fréquences

Le fait que la transformée de Fourier 3D de séquences d'images avec transparence soit composée de plans correspondant aux différentes couches (du moins dans le cas de translations) propose un moyen intuitif de les séparer. D'après (2.27), chacun des plans permet non seulement de retrouver les vitesses des différentes couches, mais aussi d'obtenir leur transformée de Fourier spatiale. Une séparation parfaite des plans pourrait ainsi permettre de reconstruire les différentes couches. Cette approche a été proposée par [Darrell 93], puis examinée de manière plus critique par Pingault dans sa thèse [Pingault 03a]. L'objection pratique essentielle est que les plans sont plus difficiles à séparer aux basses fréquences, alors que ce sont précisément les données les plus présentes et les plus critiques. Il en résulte des artefacts marqués même sur des exemples synthétiques.

En outre, il faut que le mouvement considéré soit translationnel et constant sur des dizaines d'images.

Vernon [Vernon 98] reprend la formulation en termes de décalage de phases dans l'espace de Fourier pour calculer le spectre des couches. En considérant (2.40), et si les mouvements sont bien estimés, les opérateurs de décalage Δ_1 et Δ_2 sont donnés par :

$$I_1^t(\omega_x, \omega_y) = \frac{I^t(\omega_x, \omega_y) \cdot \Delta_2(\omega_x, \omega_y) - I^{t+1}(\omega_x, \omega_y)}{\Delta_2(\omega_x, \omega_y) - \Delta_1(\omega_x, \omega_y)}, \quad (2.51)$$

$$I_2^t(\omega_x, \omega_y) = \frac{I^t(\omega_x, \omega_y) \cdot \Delta_1(\omega_x, \omega_y) - I^{t+1}(\omega_x, \omega_y)}{\Delta_1(\omega_x, \omega_y) - \Delta_2(\omega_x, \omega_y)}. \quad (2.52)$$

Un premier problème survient aux fréquences pour lesquelles Δ_1 et Δ_2 sont proches. Stuke *et al.* proposent d'interpoler les fréquences manquantes [Stuke 03b, Mota 05a]. Ils en profitent d'ailleurs pour généraliser l'approche à un nombre quelconque de couches.

Les exemples synthétiques décrits montrent des artefacts sous forme de rayures parallèles. Le problème essentiel est qu'il faut supposer le mouvement parfaitement translationnel sur une fenêtre assez grande pour calculer la transformée de Fourier, ainsi que l'a remarqué Pingault [Pingault 03a]. Cela exclut tout mouvement divergent. Ajoutons qu'aucune étude sur l'impact de légères erreurs d'estimation de mouvement n'a été menée.

2.4.2 Filtrage spatial dans l'espace direct recalé

Une alternative consiste à intégrer l'information concernant une couche sur toute la séquence. Une fois la séquence stabilisée par rapport au mouvement d'une des deux couches, on peut chercher à extraire cette couche en prenant le minimum [Szeliski 00], la moyenne [Irani 94] ou encore le médian des valeurs d'intensité des pixels. Ces résultats peuvent être améliorés en itérant ce procédé sur les séquences de résidus (séquence originale à laquelle on a ôté les estimations courantes des couches), comme proposé par Szeliski ou Hemmendorff [Hemmendorff 01].

Comme expliqué plus haut, le principe développé dans [Szeliski 00] exploite le fait que la transparence additive implique que nous observions en chaque pixel le niveau de gris de la couche stabilisée, auquel s'ajoute celui de la couche mobile. Prendre le minimum des valeurs observées sur la séquence permet donc de s'approcher de la valeur recherchée.

Les auteurs proposent un schéma itératif menant à des estimés exacts dans des conditions idéales. Leur démonstration suppose toutefois que l'image est non bruitée, que les mouvements sont entiers, et - ce sur quoi ils n'insistent pas - qu'ils sont assez réguliers pour qu'un pixel d'intensité nulle d'une couche permette de reconstruire l'autre couche.

Les approches par calcul des moyennes ou des médianes considèrent les couches non stabilisées comme des perturbations assimilables à du bruit, et estiment donc la valeur la plus probable de la couche stabilisée. Nous montrons en annexe A les résultats obtenus en appliquant cette méthode aux images d'exams cliniques que nous considérons. Ces méthodes permettent de reconstruire grossièrement l'allure des couches. Il est possible d'améliorer ces résultats en éliminant les premières valeurs estimées des couches de la séquence, puis en réitérant le calcul de la moyenne (ou du médian) [Hemendorf 01]. Il reste néanmoins souvent des artefacts se traduisant par l'apparition du fantôme d'une couche sur l'autre.

Weiss propose dans [Weiss 01] un algorithme beaucoup plus convaincant. Il cherche à calculer, à partir de n dérivées spatiales de la séquence d'origine, les n dérivées spatiales des couches, puis à les reconstruire. Nous reprenons ses notations de calcul des dérivées spatiales par convolution avec les filtres f_n .

L'auteur fait l'hypothèse que la séquence a été stabilisée par rapport à I_1 et que la couche I_2 par exemple suit la loi statistique correspondant à son histogramme⁴. Il montre que l'estimation par maximum de vraisemblance de la dérivée couche I_1 peut être donnée par le médian des dérivées de la séquence I stabilisée :

$$MAP(I_1 \star f_n) = \underset{Seq}{med}(I \star f_n). \quad (2.53)$$

Une fois ces estimateurs de différentes dérivées de la couche I_1 calculés, il s'agit de reconstruire I_1 elle-même. Weiss affirme que la meilleure reconstruction possible s'obtient comme suit :

$$\hat{I}_1 = g \star \left(\sum_n f_n^r \star MAP(I_1 \star f_n) \right) \quad (2.54)$$

avec f_n^r le filtre $f_n^r(x, y) = f_n(-x, -y)$ et g l'opérateur d'inversion tel que :

$$g \star \left(\sum_n (f_n^r \star f_n) \right) = \delta. \quad (2.55)$$

δ représentant la fonction de Dirac, qui vaut 1 en 0 et 0 sinon. En pratique, Weiss n'utilise que deux opérateurs correspondant au gradient selon x et selon y . $\sum_n f_n^r \star MAP(I_1 \star f_n)$ correspond alors au laplacien de la couche reconstruite, et g (le filtre inverse de ce laplacien) se calcule efficacement dans l'espace de Fourier.

Cette approche est bien meilleure que les précédentes. Il est plus recevable d'assimiler les mouvements du *gradient* de la couche non stabilisée à du bruit pour reconstruire le gradient de la couche stabilisée, que de considérer que la couche en mouvement elle-même se comporte comme du bruit. Cela dit, ce modèle n'en reste pas moins inexact dans

⁴Qu'il suppose laplacien, comme c'est le cas pour la plupart des images naturelles.

l'absolu. La démonstration de Weiss considère en effet que le gradient spatial de la couche I_2 suit une loi statistique laplacienne, sans aucune corrélation spatiale. Rien ne garantit donc que la reconstruction sera parfaite, mais l'expérience montre qu'elle est bien plus satisfaisante que les méthodes exposées auparavant. Nous l'avons en particulier appliqué à nos images dans l'annexe A.

Notons que l'auteur a proposé, avec Anat Levin, d'autres techniques originales [Levin 02, Levin 04a, Levin 04b] pour séparer les couches à partir d'une image uniquement. Il s'agit d'un problème très difficile, qui touche notamment à la perception humaine puisque de telles problématiques ne peuvent se résoudre qu'avec des *a priori*. Bien que les méthodes imaginées pour ce faire soient forcément très coûteuses en temps de calcul et imparfaites en ce qu'elles schématisent nécessairement la perception humaine, nous les citons rapidement car elles sont originales et intéressantes.

Dans [Levin 02], les auteurs cherchent à estimer les histogrammes fournis par les opérateurs de *dérivée* spatiale et de *détection de coin* appliqués aux deux couches. Il s'agit alors de minimiser une fonction de coût imposant aux couches extraites de se conformer à ces deux lois, et à leur somme de valoir l'image composite. Les coûts de calcul sont importants, et la minimisation se fait par transmission de croyance bouclée pour les réduire au maximum. Les exemples tendent à montrer que l'algorithme ne fonctionne que sur des images très géométriques.

Dans [Levin 04a], l'*a priori* consiste à chercher à minimiser le nombre de bords et de coins des couches extraites. L'idée sous-jacente, qui fonctionne étonnamment bien compte tenu de sa simplicité, est qu'une telle contrainte pousse à favoriser la séparation la plus simple possible, qui est aussi la plus naturelle. Ceci est mis en oeuvre en deux temps. On travaille d'abord sur des patches de l'image, que l'on cherche à décomposer à l'aide de patches issus d'une base de donnée d'images naturelles. La décomposition préférée sur l'ensemble des patches initialise alors un algorithme de propagation de croyance. Les résultats sont intéressants mais poussent à attribuer toutes les textures à une seule couche. Le schéma de minimisation est imparfait et peut aboutir à des minima locaux.

Citons enfin [Levin 04b], dans lequel l'*a priori* est donné par l'utilisateur qui clique sur l'image pour attribuer certains gradients à une couche ou à l'autre. Sous la contrainte supplémentaire que les gradients spatiaux des images suivent un histogramme laplacien, une minimisation d'abord linéaire, puis exacte basée sur une approche EM est mise en place. Les traitements présentés sont convaincants, mais nécessitent quelques centaines de clics de l'utilisateur.

Sarel et Irani considèrent la séparation de couches transparentes à travers des problématiques particulières.

Dans [Sarel 04], le problème est celui d'images pondérées de deux couches. On suppose connues deux images \mathbf{J}_1 et \mathbf{J}_2 telles que :

$$\mathbf{J}_1(\mathbf{p}) = \alpha_1(\mathbf{p})I_1(\mathbf{p}) + (1 - \alpha_1(\mathbf{p}))I_2(\mathbf{p}), \quad (2.56)$$

$$\mathbf{J}_2(\mathbf{p}) = \alpha_2(\mathbf{p})I_1(\mathbf{p}) + (1 - \alpha_2(\mathbf{p}))I_2(\mathbf{p}). \quad (2.57)$$

et il s'agit de reconstruire les couches I_1 et I_2 . Le coefficient de pondération qui varie spatialement peut modéliser des changements locaux de la physique du problème, dus

par exemple aux modifications de l'opacité du verre sur lequel se réfléchit une couche, à des saletés, etc. Une situation où l'on mesurerait deux images avec des coefficients de pondération des couches différents peut se produire quand on prend deux photos du même objet avec deux objectifs polarisateurs différents.

Le problème est résolu itérativement, en cherchant à progressivement passer l'information pertinente d'une couche à l'autre, jusqu'à ce qu'elles soient décorréélées. À chaque pas, on affine l'estimée de la couche \hat{I}_1 en lui retirant une portion (variant spatialement) $\sigma(\mathbf{p})$ de la couche \hat{I}_2 , de sorte que les deux couches soient aussi décorréélées que possible. En notant \hat{I}_1^k et \hat{I}_2^k les couches reconstruites à l'itération k :

$$\hat{I}_1^{k+1}(\mathbf{p}) = \hat{I}_1^k(\mathbf{p}) - \sigma(\mathbf{p})\hat{I}_2^k(\mathbf{p}) \quad (2.58)$$

$$\text{de sorte que } GNGC(\hat{I}_1^{k+1}, \mathbf{J}_1(\mathbf{p}) - \hat{I}_1^{k+1}) \text{ soit minimale.} \quad (2.59)$$

GNGC est une fonction de corrélation bien choisie (*Generalized Normalized Grayscale Correlation*). En pratique, on considère $\sigma(\mathbf{p})$ constant par bloc.

Les auteurs appliquent avec succès leur méthode à des images polarisées, et imaginent une extension au cas de deux couches en mouvement. Ils annulent le mouvement d'une des deux couches et la séparent en utilisant l'approche de Weiss [Weiss 01]. Il en ressort d'une part une image reconstruite de la couche stabilisée, et d'autre part la séquence privée de cette couche (qui correspond alors à la seconde couche). Comme expliqué ci-dessus, même si elle est élégante et pertinente, la méthode de Weiss n'est pas parfaite. Il subsiste souvent dans la pratique un fantôme de chacune des couches sur l'autre.

L'idée est alors d'appliquer l'algorithme "d'échange d'information" entre couches, entre la couche extraite et une image quelconque de la séquence recalée. Cela permet d'améliorer les résultats de la séparation notablement. Il n'est nécessaire pour ce faire de stabiliser qu'une seule image ; l'autre peut éventuellement comporter un mouvement complexe qui aurait été difficile à estimer. Cela dit, nous montrons en annexe A que cette méthode ne fonctionne pas bien sur les images que nous considérons.

Sarel et Irani ont également étendu leur approche à une situation de deux mouvements arbitrairement complexes, sous la condition que l'un des deux soit périodique. La contribution essentielle de leur article [Sarel 05] est une méthode de recalage adaptée à cette configuration. L'unité de base de leur algorithme n'est pas l'image mais un segment de la séquence P (comptant jusqu'à quelques dizaines d'images), correspondant grossièrement à une période du mouvement.

Une fois un tel segment d'intérêt P défini, les auteurs cherchent à aligner toutes les images qui se correspondent du fait du mouvement périodique. Pour ce faire, il est appliqué au segment de référence une transformation affine 3D globale pour l'aligner sur les segments P_t de la séquence (en maximisant une mesure de corrélation). Un seuillage sur la valeur de la corrélation obtenue permet de déterminer si le recalage est pertinent ou non.

Dans le cas où P et P_t se correspondent vraiment, leur recalage est affiné localement en autorisant de faibles décalages spatiaux et temporels. Il ne reste alors qu'à appliquer la méthode de Weiss, puis "l'échange d'information" entre couches, entre les images mises en correspondance pour reconstruire les deux couches recherchées.

L'accent est mis dans l'article sur la mise en correspondance des segments plus que sur la

séparation des couches. La méthode semble particulièrement solide, puisqu'elle s'appuie non pas sur une image mais sur quelques dizaines d'images consécutives. Elle pourrait s'appliquer aux mouvements anatomiques qui sont périodiques (coeur, diaphragme, poumons, etc).

Par contre, les temps de calcul sont probablement considérables, et les exemples donnés sont très particuliers. S'ils donnent lieu à une séparation convaincante, ils présentent des situations où le mouvement non périodique est hautement incohérent (et parfois peu contrasté), ce qui permet de le considérer aisément comme une perturbation. On peut se demander par contre comment se comporterait l'algorithme si les deux mouvements étaient à peu près également structurés.

Ces deux dernières méthodes présentent probablement les meilleurs résultats de séparation des couches à ce jour. Cependant, elles sont très lourdes en temps de calcul, et nous ne savons rien de leur comportement face au bruit.

2.4.3 Résolution directe du système

Une autre voie prometteuse est esquissée dans [Szeliski 00]. Les auteurs rassemblent les contraintes imposées par la transparence en chaque pixel de l'image composite dans un système linéaire de large dimension. Notant \mathbf{J}_i un vecteur colonne correspondant à la couche transparente i , $\mathbf{J}(t)$ un vecteur colonne correspondant à l'image I à la date t , les équations de transparence

$$\sum_i \mathbf{I}_i(\mathbf{p} + \mathbf{w}) = \mathbf{I}(\mathbf{p}, t) \quad (2.60)$$

peuvent s'écrire

$$\sum_i \mathbf{W}_t \mathbf{J}_i = \mathbf{J}(t). \quad (2.61)$$

\mathbf{W}_t , la matrice qui modélise le déplacement des couches, se calcule à partir de leur mouvement et du type d'interpolation choisi. C'est une matrice de très large dimension, mais qui est creuse car chacune de ses lignes ne représente l'interpolation que d'un seul point. Les auteurs proposent de résoudre le problème des moindres carrés contraints

$$\min_{\mathbf{J}_i, i \in [1, n]} \sum_t \left\| \sum_i \mathbf{W}_t \mathbf{J}_i - \mathbf{J}(t) \right\|^2. \quad (2.62)$$

Ils proposent une implantation pratique de cette méthode de séparation des couches, mais ne s'étendent pas sur ses propriétés en terme de temps de calcul. Cette méthode est prometteuse, mais l'étude du comportement de l'estimateur proposé est délicate.

Les auteurs mettent notamment en garde contre l'impact du bruit, de l'interpolation, et surtout du type de mouvement suivi par les couches. La méthode proposée ne permet en effet qu'une séparation à un offset près, ce qui peut donner des résultats désastreux quand les couches peuvent être segmentées en des régions ayant des trajectoires disjointes. Une même couche y aura des offsets différents selon la région considérée.

2.4.4 Séparation des couches appliquée aux images fluoroscopiques

Nous avons implanté et testé les trois méthodes automatiques les plus matures de séparation de couches sur des séquences d'images cliniques dans l'annexe A. Il s'agit :

- du calcul de médian temporel sur la séquence stabilisée par rapport à une couche ;
- de la méthode de [Weiss 01] (le médian temporel est appliqué aux gradients) ;
- de la méthode d'échange d'information des couches [Sarel 04].

Nous y concluons que l'utilisation du médian temporel sur les séquences stabilisées donne des résultats esthétiquement intéressants, qui sont meilleurs lorsque le médian est pris sur des images de gradient.

Cela dit, ces méthodes ne sont applicables en pratique qu'à la couche statique, car elles sont très sensibles à la qualité de la compensation de mouvement de la couche considérée. En outre, si la couche séparée fournit une bonne approximation visuelle de la couche réelle, elle ne vérifie pas précisément la physique de la transparence additive, de sorte que la séquence d'images résiduelle est inutilisable.

Nous avons alors essayé d'améliorer la séparation par la méthode d'échange d'information, mais sans succès. Quand elle est utilisée au niveau de l'image seulement, le transfert d'information est si faible que les couches extraites ne sont quasiment pas modifiées, alors que quand elle est appliquée de manière plus locale, des problèmes de stabilité apparaissent.

Au final donc, aucune des méthodes actuelles de séparation des couches n'est applicable à notre problématique, que ce soit pour estimer les mouvements transparents ou pour débruiter les images transparentes dans le sens du mouvement. Soulignons enfin que les méthodes présentées impliquent un volume de calcul important puisque le médian doit être pris sur une dizaine d'images au moins pour donner des résultats intéressants.

2.5 Conclusion

Nous avons présenté dans ce chapitre l'état de l'art sur l'estimation de mouvements additivement transparents. Si des estimateurs robustes monocouche peuvent fournir dans quelques cas de bonnes approximations des champs de vitesses, il devient vite évident qu'une approche générale du problème nécessite une modélisation explicite de la transparence.

Shizawa et Mase ont été pionniers dans ce domaine. Ils ont proposé deux extensions de l'équation du flot optique dans l'espace des fréquences et dans l'espace direct qui prenaient explicitement en compte la transparence. Les mises en œuvre algorithmiques qu'ils proposent, basées sur des diagonalisations tensorielles, sont moins convaincantes.

Une première grande classe de méthodes d'estimation se place dans le cadre fréquentiel. Si les mouvements des couches sont translationnels et constants sur un nombre important d'images, la présence de ces couches se traduit dans l'espace de Fourier $3D$ par la concentration des spectres dans des plans dont les coefficients directeurs permettent de calculer les vitesses. Les multiples contributions effectuées dans ce cadre ne sont cependant pas applicables sur des images cliniques car les mouvements anatomiques sont loins d'être constants sur des plages temporelles suffisantes.

La deuxième famille d'estimateurs se base sur une extension de la conservation de l'intensité à la transparence. Sous l'hypothèse que les mouvements sont constants entre trois dates successives, et qu'ils varient peu spatialement, il est en effet possible d'exhiber une

équation simple liant vitesses à estimer et niveaux de gris de trois images successives de la séquence.

La première hypothèse porte sur un horizon temporel beaucoup plus raisonnable que les approches fréquentielles, et nous en discuterons plus loin. L'impact d'un écart de la seconde contrainte, que nous sommes les seuls à avoir soulevé, doit être étudié avec soin. Il dépend de la divergence des mouvements ainsi que du contraste des images considérées. Il existe par ailleurs des équivalents à l'équation de la conservation de l'intensité généralisée pour n couches, mais il faut alors supposer les mouvements constants sur $n + 1$ images consécutives.

Nous avons classé les contributions exploitant cette équation selon qu'elles introduisent un modèle des champs de vitesse ou non. Nous retrouvons dans la première catégorie des cas des vitesses translationnelles par bloc, ou la modélisation par B-splines. Les estimateurs de champs denses, qu'ils s'appuient sur une minimisation d'une fonctionnelle avec un terme de lissage, ou relèvent d'un formalisme markovien, se rangent dans la seconde catégorie. Nous avons également cité une approche intermédiaire commençant par estimer des champs denses avant de les segmenter avec prise en compte de modèles affines de mouvement.

Enfin, nous avons fait une incursion dans la problématique de la séparation des couches. Les approches les plus intuitives (séparation des plans dans l'espace de Fourier $3D$, intégration de l'information sur des séquences d'images recalées) sont sujettes à des artefacts. Des méthodes plus novatrices, basées sur un filtrage portant sur les dérivées spatiales de l'image ou sur "l'échange d'information" entre couches donnent des résultats plus acceptables. Cela dit, nous montrons dans l'annexe A que le problème n'est pas résolu de manière satisfaisante pour les images que nous considérons de sorte que nous chercherons plutôt à le contourner dans cette thèse.

Chapitre 3

Estimation du mouvement en situation de bi-transparence pour les examens par rayons X

Nous ne retenons des pistes d'estimation du mouvement transparent présentées dans le chapitre 2 que celles qui s'appuient sur l'équation de contrainte du mouvement transparent dans l'espace direct (2.39). Nous avons en effet expliqué d'une part que seule une prise en compte explicite de la transparence permettait de traiter les phénomènes de transparence totale dus aux rayons X, et d'autre part qu'une estimation de mouvement dans le domaine de Fourier impliquait une constance des mouvements dans le temps incompatible avec le rythme cardiaque.

Nous nous plaçons dans ce chapitre dans le cas d'une transparence à deux couches présente sur toute l'image, ce que nous appelons la *bi-transparence*. Le cas général sera traité dans les chapitres 4 et 5 à travers l'introduction du concept de *bi-transparence distribuée*. Pour traiter les images difficiles issues des examens fluoroscopiques (relativement bruitées et peu contrastées), nous proposons de contraindre le problème via l'introduction de champs de mouvement paramétriques. Le bien-fondé de cette approche est discuté au paragraphe 3.1. Le problème se ramène alors à la minimisation d'une fonctionnelle que nous résolvons par une méthode incrémentale de type Gauss-Newton, détaillée au paragraphe 3.2. Une version robuste de cette résolution est proposée au paragraphe 3.3.

Ce type de résolution étant particulièrement sensible à l'initialisation, surtout sur des données bruitées, nous détaillons la méthode d'initialisation choisie en section 3.4. Nous concluons en donnant une mesure quantitative des performances de notre estimateur à l'aide des images réalistes introduites dans le chapitre 1, et présentons des résultats sur des séquences cliniques réelles diagnostiques et interventionnelles.

3.1 Équation de contrainte du mouvement transparent avec modèles paramétriques

Nous considérons dans ce chapitre une séquence de trois images en bi-transparence, et nous cherchons à estimer les mouvements des deux couches présentes. Dans le cadre qui nous intéresse, ces deux couches correspondent au mouvement de deux portions d'organes.

Nous avons en effet précisé dans notre introduction que nous nous focalisons uniquement sur les mouvements des organes, en excluant les outils interventionnels¹.

Pour estimer les mouvements des deux couches organiques, nous nous appuyons sur (2.39), que nous rappelons ici :

$$\begin{aligned} r(\mathbf{p}, \mathbf{w}_1(\mathbf{p}), \mathbf{w}_2(\mathbf{p})) &= I(\mathbf{p}, t + 1) - I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t) \\ &\quad - I(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t) + I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}) + \mathbf{w}_2(\mathbf{p}), t - 1) = 0. \end{aligned} \quad (3.1)$$

Estimer les champs des vitesses \mathbf{w}_1 et \mathbf{w}_2 sur toute l'image revient à minimiser :

$$J(\mathbf{w}_1, \mathbf{w}_2) = \sum_{\mathbf{p} \in \mathfrak{S}} r(\mathbf{p}, \mathbf{w}_1(\mathbf{p}), \mathbf{w}_2(\mathbf{p}))^2. \quad (3.2)$$

Les différentes méthodes d'estimation du mouvement transparent basées sur cette équation varient essentiellement par les hypothèses qu'elles font sur les mouvements recherchés. Plus ces dernières sont flexibles, plus elles permettent d'estimer précisément des mouvements complexes, mais plus l'algorithme est sensible au bruit et exigeant en temps de calcul. Il faut donc proposer un compromis entre la généralité de la classe de mouvements traités d'une part, et la complexité algorithmique, la robustesse au bruit et la sensibilité au réglage des paramètres d'autre part.

Nous pouvons classer à une extrémité de ce spectre les calculs de champs denses [Stuke 03b, Stuke 04], et à l'autre l'approximation de mouvements translationnels par blocs [Pingault 02b, Stuke 03a]. L'approche de décomposition des champs de vitesse sur des bases de B-splines [Pingault 03b] peut être vue comme intermédiaire : elle permet de modéliser des mouvements complexes, tout en étant moins coûteuse que les méthodes denses.

Dans notre cas particulier, nous sommes en mesure de contraindre le mouvement à estimer. Les organes ont en effet des mouvements assez réguliers, que ce soient les battements du coeur, la dilatation des poumons ou la translation du diaphragme. Comme notre estimateur se doit d'être particulièrement robuste au bruit des images fluoroscopiques, et assez rapide pour qu'une implantation future en temps réel soit envisageable, nous nous devons d'utiliser toutes les contraintes pertinentes afin de délimiter aussi précisément que possible le problème traité.

C'est pourquoi nous proposons de représenter les champs de vitesses par des modèles polynômiaux 2D. Nous choisissons les modèles de mouvement affines, puisqu'ils permettent de représenter les mouvements de translation, de dilatation/contraction, de rotation ainsi que de cisaillement, tout en ne mettant en jeu que 6 paramètres par couche.

En toute rigueur, les mouvements des organes sont plus complexes, ne serait-ce que parce qu'ils sont la projection 2D de mouvements 3D. Il est donc légitime de s'interroger sur la validité de cette hypothèse. Mais les exemples cliniques que nous présentons dans les chapitre 3, 4 et 5 montrent que ce type de modélisation, allié à la méthode de segmentation automatique des couches présentée dans le chapitre 5, permet des estimés de mouvement fiables en situation réelle. Par ailleurs, il est délicat d'estimer le mouvement des organes plus finement sans *a priori* sur l'anatomie considérée [Paragios 98]. Notons d'ailleurs que notre méthode peut aisément s'étendre à des modèles plus compliqués, tels que les champs

¹L'estimation de leur mouvement nous semble en effet nécessiter la mise au point de méthodes spécifiques. Une approche générale ne pourra à notre avis capter leurs déplacements qu'à travers celui des organes auxquels ils sont liés.

de mouvement quadratiques, si besoin est.

Ainsi, le vecteur mouvement de la couche k ($k = 1, 2$) au point $\mathbf{p} = (x, y)$ est représenté par \mathbf{w}_{θ_k} :

$$\begin{cases} u_{\theta_k}(x, y) = a_{1,k} + a_{2,k} \cdot x + a_{3,k} \cdot y \\ v_{\theta_k}(x, y) = a_{4,k} + a_{5,k} \cdot x + a_{6,k} \cdot y. \end{cases} \quad (3.3)$$

Nous regroupons les paramètres de mouvement dans le vecteur $\theta_k = (a_{1,k}, \dots, a_{6,k})$, et posons $\Theta = (\theta_1, \theta_2)$. Sous l'hypothèse que les mouvements présents dans l'image sont effectivement globalement affines², le problème revient à estimer les valeurs des paramètres θ_1 et θ_2 permettant de minimiser :

$$J(\Theta) = \sum_{\mathbf{p} \in \mathfrak{S}} r(x, y, \theta_1, \theta_2)^2 \quad (3.4)$$

$$\begin{aligned} \text{avec } r(\mathbf{p}, \theta_1, \theta_2) &= I(\mathbf{p}, t + 1) - I(\mathbf{p} + \mathbf{w}_{\theta_1}, t) \\ &\quad - I(\mathbf{p} + \mathbf{w}_{\theta_2}, t) + I(\mathbf{p} + \mathbf{w}_{\theta_1} + \mathbf{w}_{\theta_2}, t - 1) \end{aligned} \quad (3.5)$$

les composantes des vitesses $u_{\theta_1}, v_{\theta_1}, u_{\theta_2}$ et v_{θ_2} étant liées aux paramètres de mouvement par les deux relations (3.3).

Notons que ce critère est parfaitement symétrique en θ_1 et θ_2 (puisque rien ne nous impose d'attribuer l'étiquette 1 à un champ plutôt qu'à un autre). L'énergie $J(\Theta)$ a en particulier deux minima symétriques.

Nous avons représenté sur la figure 3.1 le tracé de l'énergie J en fonction des paramètres de mouvement, pour une séquence d'images transparentes dont les deux couches sont en translation suivant l'axe des abscisses. Nous avons restreint l'espace de représentation aux deux composantes translationnelles $a_{1,1}$ et $a_{1,2}$, toutes les autres étant fixées à la valeur 0. Nous avons représenté deux cas : l'un où les composantes recherchées sont clairement distinctes et l'autre où elles sont assez proches.

Nous remarquons que les minima sont bien marqués dans le premier cas, mais moins nets dans le second cas. S'il est aisé d'estimer un des deux mouvements, l'estimation du deuxième mouvement est alors problématique. Dans la configuration présentée à droite de la figure 3.1, un observateur ne percevrait d'ailleurs qu'une situation mono-couche en regardant la séquence tant les mouvements sont proches, ce qui relativise ce problème.

3.2 Estimation multirésolution du mouvement transparent

L'expression (3.4) étant non-linéaire, sa minimisation est à considérer avec attention. Si les mouvements recherchés étaient de faible amplitude, nous pourrions linéariser (3.4) comme détaillé dans la sous-section suivante, et utiliser une méthode efficace de minimisation itérative pour les estimer. Comme de larges mouvements peuvent aussi apparaître en pratique, nous introduisons un schéma multirésolution incrémental utilisant des pyramides de B-splines.

²Qui s'ajoute, rappelons-le, à l'hypothèse du mouvement constant sur 3 images successives, et à celle, moins contraignante, de mouvement localement translationnel.

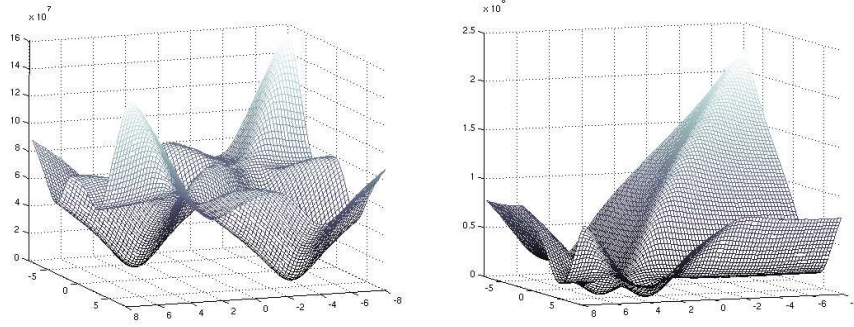


FIG. 3.1: Tracé de l'énergie J en fonction des paramètres de mouvement pour une séquence d'images transparentes dont les deux couches sont en translation suivant l'axe des abscisses. L'espace de représentation est restreint aux deux composantes translationnelles $a_{1,1}$ et $a_{1,2}$, toutes les autres étant fixées à la valeur 0. À gauche, les deux composantes sont clairement distinctes, à droite elles sont proches.

3.2.1 Linéarisation et résolution

Une remarque nous est utile pour les calculs qui vont suivre. Les vecteurs \mathbf{w}_{θ_i} ($i = 1, 2$) de déplacement sont liés de manière linéaire aux paramètres de mouvement θ_i :

$$\mathbf{w}_{\theta_i} = \mathbf{B}(\mathbf{p})\theta_i \quad (3.6)$$

avec la matrice $\mathbf{B}(\mathbf{p})$ qui vaut :

$$\mathbf{B}(\mathbf{p}) = \begin{bmatrix} 1 & x & y & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x & y \end{bmatrix} \quad (3.7)$$

Une conséquence intéressante pour les linéarisations à venir est que :

$$\mathbf{w}_{\theta_i + \delta\theta_i} = \mathbf{w}_{\theta_i} + \mathbf{w}_{\delta\theta_i} \quad (3.8)$$

$$= \mathbf{w}_{\theta_i} + \delta\mathbf{w}_{\theta_i} \quad (3.9)$$

en notant $\mathbf{w}_{\delta\theta_i} = \delta\mathbf{w}_{\theta_i}$.

3.2.1.1 Au premier ordre

Les calculs qui vont suivre sont développés dans l'annexe B. Nous souhaitons minimiser J en nous basant sur une linéarisation de r au premier ordre :

$$J(\Theta + \delta\Theta) = \sum_{\mathbf{p} \in \mathfrak{S}} \left(r(\mathbf{p}, \Theta + \delta\Theta) \right)^2 \quad (3.10)$$

Or, r vaut d'après (3.1)

$$\begin{aligned} r(\mathbf{p}, \Theta + \delta\Theta) &= I(\mathbf{p}, t + 1) - I(\mathbf{p} + \mathbf{w}_{\theta_1 + \delta\theta_1}(\mathbf{p}), t) - I(\mathbf{p} + \mathbf{w}_{\theta_2 + \delta\theta_2}(\mathbf{p}), t) \\ &\quad + I(\mathbf{p} + \mathbf{w}_{\theta_1 + \delta\theta_1}(\mathbf{p}) + \mathbf{w}_{\theta_2 + \delta\theta_2}(\mathbf{p}), t - 1) \end{aligned} \quad (3.11)$$

$$\begin{aligned} &= I(\mathbf{p}, t + 1) - I(\mathbf{p} + \mathbf{w}_{\theta_1}(\mathbf{p}) + \delta\mathbf{w}_{\theta_1}(\mathbf{p}), t) - I(\mathbf{p} + \mathbf{w}_{\theta_2}(\mathbf{p}) + \delta\mathbf{w}_{\theta_2}(\mathbf{p}), t) \\ &\quad + I(\mathbf{p} + \mathbf{w}_{\theta_1}(\mathbf{p}) + \delta\mathbf{w}_{\theta_1}(\mathbf{p}) + \mathbf{w}_{\theta_2}(\mathbf{p}) + \delta\mathbf{w}_{\theta_2}(\mathbf{p}), t - 1) \end{aligned} \quad (3.12)$$

d'après (3.9). En utilisant une formule de Taylor :

$$\begin{aligned}
 r(\mathbf{p}, \Theta + \delta\Theta) &= I(\mathbf{p}, t + 1) - I(\mathbf{p} + \mathbf{w}_{\theta_1}(\mathbf{p}), t) - \delta\mathbf{w}_{\theta_1} \nabla I(\mathbf{p} + \mathbf{w}_{\theta_1}(\mathbf{p}), t) \\
 &\quad - I(\mathbf{p} + \mathbf{w}_{\theta_2}(\mathbf{p}), t) - \delta\mathbf{w}_{\theta_2} \nabla I(\mathbf{p} + \mathbf{w}_{\theta_2}(\mathbf{p}), t) \\
 &\quad + I(\mathbf{p} + \mathbf{w}_{\theta_1}(\mathbf{p}) + \mathbf{w}_{\theta_2}(\mathbf{p}), t - 1) \\
 &\quad + (\delta\mathbf{w}_{\theta_1} + \delta\mathbf{w}_{\theta_2}) \nabla I(\mathbf{p} + \mathbf{w}_{\theta_1}(\mathbf{p}) + \mathbf{w}_{\theta_2}(\mathbf{p}), t - 1) \tag{3.13}
 \end{aligned}$$

$$\begin{aligned}
 &= r(\mathbf{p}, \Theta) \\
 &\quad + (\nabla I(\mathbf{p} + \mathbf{w}_{\theta_1}(\mathbf{p}) + \mathbf{w}_{\theta_2}(\mathbf{p}), t - 1) \\
 &\quad \quad - \nabla I(\mathbf{p} + \mathbf{w}_{\theta_1}(\mathbf{p}), t)) \delta\mathbf{w}_{\theta_1}(\mathbf{p}) \\
 &\quad + (\nabla I(\mathbf{p} + \mathbf{w}_{\theta_1}(\mathbf{p}) + \mathbf{w}_{\theta_2}(\mathbf{p}), t - 1) \\
 &\quad \quad - \nabla I(\mathbf{p} + \mathbf{w}_{\theta_2}(\mathbf{p}), t)) \delta\mathbf{w}_{\theta_2}(\mathbf{p}) \tag{3.14}
 \end{aligned}$$

$$\begin{aligned}
 &= r(\mathbf{p}, \Theta) \\
 &\quad + (\nabla I(\mathbf{p} + \mathbf{w}_{\theta_1}(\mathbf{p}) + \mathbf{w}_{\theta_2}(\mathbf{p}), t - 1) \\
 &\quad \quad - \nabla I(\mathbf{p} + \mathbf{w}_{\theta_1}(\mathbf{p}), t)) \mathbf{B}(\mathbf{p}) \delta\theta_1 \\
 &\quad + (\nabla I(\mathbf{p} + \mathbf{w}_{\theta_1}(\mathbf{p}) + \mathbf{w}_{\theta_2}(\mathbf{p}), t - 1) \\
 &\quad \quad - \nabla I(\mathbf{p} + \mathbf{w}_{\theta_2}(\mathbf{p}), t)) \mathbf{B}(\mathbf{p}) \delta\theta_2. \tag{3.15}
 \end{aligned}$$

Ce que nous notons :

$$r(\mathbf{p}, \Theta + \delta\Theta) = r(\mathbf{p}, \Theta) + \nabla_{\theta} r(\mathbf{p}, \Theta) \delta\Theta. \tag{3.16}$$

L'énergie J à minimiser devient :

$$J_1(\Theta + \delta\Theta) = \sum_{\mathbf{p} \in \mathfrak{S}} \left(r(\mathbf{p}, \Theta) + \delta\Theta \nabla_{\Theta} r(\mathbf{p}, \Theta) \right)^2. \tag{3.17}$$

Nous souhaitons minimiser cette expression par rapport à $\delta\Theta$:

$$\begin{aligned}
 \frac{\partial J_1(\Theta + \delta\Theta)}{\partial \delta\Theta} &= 2. \sum_{\mathbf{p} \in \mathfrak{S}} \frac{\partial (r(\mathbf{p}, \Theta) + \delta\Theta \nabla_{\Theta} r(\mathbf{p}, \Theta))}{\partial \delta\Theta} \\
 &\quad \cdot (r(\mathbf{p}, \Theta) + \delta\Theta \nabla_{\Theta} r(\mathbf{p}, \Theta)) \tag{3.18}
 \end{aligned}$$

$$= 2. \sum_{\mathbf{p} \in \mathfrak{S}} \nabla_{\Theta} r(\mathbf{p}, \Theta) (r(\mathbf{p}, \Theta) + \delta\Theta \nabla_{\Theta} r(\mathbf{p}, \Theta)) \tag{3.19}$$

$$\begin{aligned}
 &= \delta\Theta \left(2. \sum_{\mathbf{p} \in \mathfrak{S}} (\nabla_{\Theta} r(\mathbf{p}, \Theta))^2 \right) \\
 &\quad + \left(2. \sum_{\mathbf{p} \in \mathfrak{S}} r(\mathbf{p}, \Theta) \nabla_{\Theta} r(\mathbf{p}, \Theta) \right) \tag{3.20}
 \end{aligned}$$

$$= 0. \tag{3.21}$$

D'où :

$$\delta\Theta = - \left(\sum_{\mathbf{p} \in \mathfrak{S}} (\nabla_{\Theta} r(\mathbf{p}, \Theta))^2 \right)^{-1} \cdot \left(\sum_{\mathbf{p} \in \mathfrak{S}} r(\mathbf{p}, \Theta) \nabla_{\Theta} r(\mathbf{p}, \Theta) \right). \tag{3.22}$$

On peut donc calculer $\delta\Theta$ à l'aide d'une inversion matricielle.

3.2.1.2 Au second ordre

L'expression (3.1) linéarisée au second ordre autour de Θ s'écrit, en utilisant de la même façon un développement limité via la formule de Taylor :

$$r_2(\mathbf{p}, \Theta + \delta\Theta) = r(\mathbf{p}, \Theta) + \delta\Theta \nabla_{\Theta} r(\mathbf{p}, \Theta) + \frac{1}{2} \delta\Theta^T H_{\Theta}(r(\mathbf{p}, \Theta)) \delta\Theta \quad (3.23)$$

où $H_{\Theta}(r(\mathbf{p}, \Theta))$ représente le hessien de r , dont l'expression exacte en fonction des dérivées spatiales des images recalées est donnée en annexe B.

Dans ces conditions, nous sommes en mesure de calculer analytiquement la fonction suivante, qu'il convient de minimiser :

$$J_2(\Theta + \delta\Theta) = \sum_{\mathbf{p} \in \mathfrak{S}} \left(r(\mathbf{p}, \Theta + \delta\Theta \nabla_{\Theta} r(\mathbf{p}, \Theta) + \frac{1}{2} \delta\Theta^T H_{\Theta}(r(\mathbf{p}, \Theta)) \delta\Theta) \right)^2. \quad (3.24)$$

Par souci de clarté, nous introduisons la notation :

$$K(\delta\Theta) = J_2(\Theta + \delta\Theta). \quad (3.25)$$

Cette écriture met en évidence le fait que Θ est donné dans cette sous-section, la discussion qui suit cherchant à établir quel $\delta\Theta$ minimise l'expression (3.25).

Nous représentons un exemple de cette énergie $K(\delta\Theta)$ à la figure 3.2, dans le cas d'une séquence d'images transparentes dont les couches sont en translation suivant l'axe des abscisses. L'espace de représentation est restreint aux deux composantes translationnelles $\delta a_{1,1}$ et $\delta a_{1,2}$. Il est bien plus facile de minimiser une énergie régulière de ce type que de travailler sur l'énergie J directement (Fig.3.1).

Comme le montre (3.24), K est un polynôme de degré 4, que nous ne pouvons pas minimiser directement comme en 3.2.1.1, puisqu'une dérivation mènerait à un polynôme de degré 3 à annuler (dans un espace de dimension 12!). Nous avons donc recours à la méthode de minimisation du *gradient conjugué*, qui fonctionne en deux temps.

Il s'agit de choisir une direction de descente permettant de se rapprocher du minimum, puis de fixer le pas de descente de sorte à minimiser (3.25) au mieux. Le procédé est itéré un certain nombre de fois, et est stoppé quand les estimés successifs sont très proches (Fig.3.3).

Plus précisément, partant de l'estimé du paramètre à l'itération précédente $\delta\Theta^k$,

- on choisit une direction de descente $d(\delta\Theta^k)$ de l'énergie $K(\cdot)$ en $\delta\Theta^k$. La direction la plus intuitive serait le gradient de K en $\delta\Theta^k$: $\nabla_{\delta\Theta} K(\delta\Theta^k)$. Ce choix ignore cependant l'historique de la minimisation, et amène parfois à mettre à jour l'estimé du minimum dans un sous-espace correspondant à la direction de plus forte pente, en ignorant le reste de l'espace de recherche.

La formule de Polak et Ribière [Polak 71] permet d'éviter cet écueil en construisant $d(\delta\Theta^k)$ en fonction de la direction choisie à l'itération précédente $d(\delta\Theta^{k-1})$ et du gradient de la fonction $K(\delta\Theta)$ aux itérations k et $k+1$:

$$d(\delta\Theta^k) = -\nabla_{\delta\Theta} K(\delta\Theta^k) + \gamma d(\delta\Theta^{k-1}) \quad (3.26)$$

avec :

$$\gamma = \frac{\left(\nabla_{\delta\Theta} K(\delta\Theta^k) - \nabla_{\delta\Theta} K(\delta\Theta^{k-1}) \right)^T \nabla_{\delta\Theta} K(\delta\Theta^k)}{\nabla_{\delta\Theta} K(\delta\Theta^k)^T \nabla_{\delta\Theta} K(\delta\Theta^k)} \quad (3.27)$$

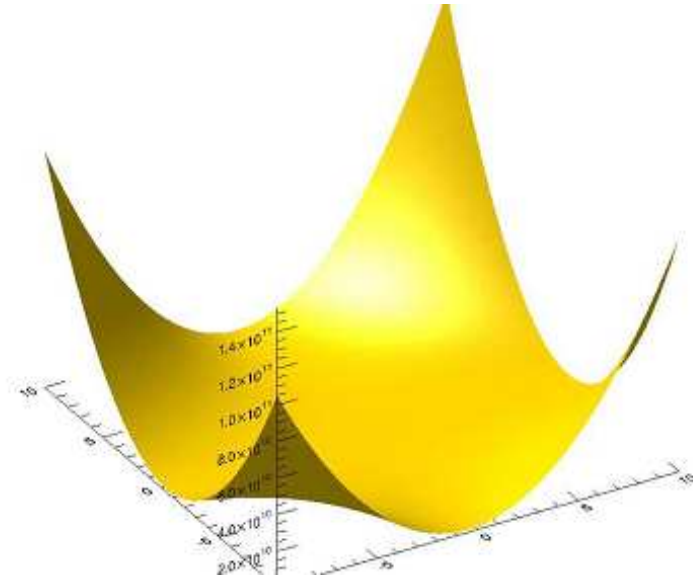


FIG. 3.2: Tracé de l'énergie $K(\delta\theta)$ en fonction des paramètres de mouvement pour une séquence d'images transparentes dont les couches sont en translation suivant l'axe des abscisses. L'espace de représentation est restreint aux deux composantes translationnelles $\delta a_{1,1}$ et $\delta a_{1,2}$, toutes les autres étant fixées à 0.

En fait, nous expliquons en annexe B qu'il faut prendre la précaution d'"homogénéiser" l'espace des $\delta\Theta$ avant d'effectuer ces calculs, ce qui mène à des formules légèrement différentes.

- une fois la direction de descente choisie, il faut calculer le pas α^k tel que

$$\delta\Theta^{k+1} = \delta\Theta^k + \alpha^k \cdot d(\delta\Theta^k) \quad (3.28)$$

minimise (3.24). Alors que de nombreuses méthodes classiques de minimisation choisissent ce pas par dichotomie, il est possible de faire mieux pour le type de fonction d'énergie que nous minimisons. En substituant $\delta\Theta^{k+1}$ par son expression (3.28) dans l'expression de $K(\delta\Theta)$ (3.24), nous remarquons qu'il s'agit d'un polynôme de degré 4 dans l'espace de dimension 1 des α^k . Il est ainsi possible de déterminer exactement la valeur optimale du pas en annulant sa dérivée par rapport à α^k , un polynôme (1D) de degré trois³. Les calculs complets sont menés dans l'annexe B.

La minimisation selon (3.21) garantit l'absence de minima locaux. En revanche, la linéarisation au second ordre a l'avantage de mieux représenter l'allure locale de $J(\Theta)$ par nature, et ce faisant elle permet de s'approcher avec précision des minima de J même lorsqu'ils sont relativement éloignés du point de linéarisation. Pour atteindre le même minimum, la première méthode devra avoir recours à plusieurs linéarisations successives.

Il s'avère d'après nos tests que la linéarisation au second ordre permet d'estimer les mou-

³En fait, il s'avère moins rapide d'utiliser les formules analytiques des zéros des polynômes de degré 3 que de les rechercher par la méthode de Newton, ce qui ne change pas le fait que nous calculons de manière exacte le pas optimal recherché.

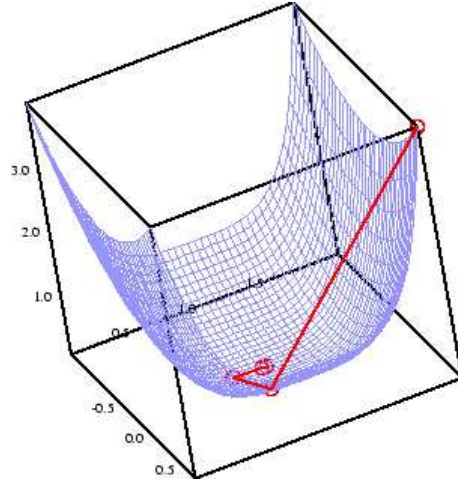


FIG. 3.3: Illustration de la minimisation itérative de fonctionnelle par la méthode des gradients conjugués.

vements nettement plus rapidement que la première méthode, à précision égale. Ceci sera détaillé au paragraphe 3.5.

3.2.2 Formalisme multirésolution

Les méthodes exposées précédemment reposent sur des *linéarisations* de l'énergie à minimiser, et ne sont donc valides que pour des petits mouvements. Comme les images cliniques présentent de larges mouvements (de l'ordre de 8 pixels à certaines dates du cycle cardiaque sous certaines angulations), nous introduisons un schéma multirésolution incrémental exploitant des pyramides de splines [Unser 93] des trois images consécutives⁴ (Fig.3.4).

Au niveau de résolution le plus grossier L , les mouvements sont suffisamment petits pour que l'on puisse appliquer une des méthodes de minimisation par linéarisation présentées dans la sous-section précédente. Il est d'ailleurs possible d'itérer, en effectuant quelques linéarisations successivement à un même niveau. Notons les estimés finaux au niveau L $(\hat{\theta}_1^L, \hat{\theta}_2^L)$.

Au niveau $L-1$, nous initialisons $(\theta_1^{L-1}, \theta_2^{L-1})$ avec $(\tilde{\theta}_1^{L-1}, \tilde{\theta}_2^{L-1})$, où $\tilde{a}_{i,k}^{L-1} = 2\hat{a}_{i,k}^L$ ($i = 1, 4$) et $\tilde{a}_{j,l}^{L-1} = \hat{a}_{j,l}^L$ ($j = 2, 3, 5, 6$). Il ne reste alors qu'à appliquer l'une des deux méthodes de la partie précédente à $J(\Theta^{L-1})$ en écrivant $\Theta^{L-1} = \tilde{\Theta}^{L-1} + \delta\Theta^{L-1}$.

Cette méthode de Gauss-Newton incrémentale est itérée à travers les niveaux de résolution successifs jusqu'au plus fin d'entre eux. Outre le fait qu'elle permet d'appréhender de larges mouvements, cette méthode a aussi l'avantage d'être rapide et relativement robuste au bruit. En effet, le bruit est fortement filtré aux bas niveaux de résolution lors de la construction de la pyramide, ce qui se traduit par des fonctionnelles plus régulières, dont les minima locaux sont bien moins marqués, comme illustré à la Fig.3.5.

⁴Nous avons choisi la construction de pyramides de splines par souci de cohérence avec les interpolations utilisées par ailleurs.

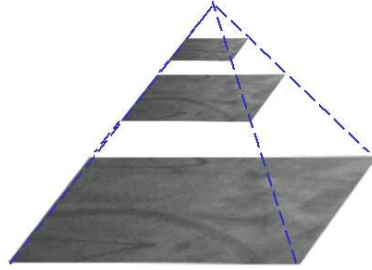


FIG. 3.4: Pyramide à plusieurs résolutions d'une image.

3.3 Estimation robuste du mouvement transparent

3.3.1 Intérêt d'une estimation robuste

Dans la section précédente, pour estimer les deux champs affines de mouvement, nous avons choisi d'imposer la contrainte de mouvement transparent (3.1) sur toute l'image à travers une somme quadratique des résidus sur l'ensemble des pixels. Si toutes les données respectent le modèle à un bruit près⁵, c'est effectivement le critère optimal qu'il s'agit de minimiser.

En présence de pixels s'écartant du modèle, appelés *outliers*, l'utilisation d'un critère quadratique entraîne rapidement un biais d'estimation. Dans notre cas, il pourra s'agir de pixels correspondant à des outils interventionnels ayant un mouvement indépendant de celui des organes (un guide ou un cathéter déplacés par le médecin par exemple), ou encore d'organes ayant un mouvement parasite sur une petite portion de l'image. Un critère quadratique donnerait une importance considérable à ces pixels. Les champs estimés finalement seraient alors notablement erronés.

Nous voudrions au contraire ignorer ces outliers pour ne nous baser que sur les informations de mouvement pertinentes.

3.3.2 La fonction robuste choisie

La statistique robuste [Hubert 81] permet de tolérer de grandes déviations au modèle de donnée à travers l'introduction de fonctions de coût pénalisant moins les grands résidus que la fonction quadratique.

Il existe une large famille de fonctions robustes, adaptées à des problématiques différentes [Black 96b] : Cauchy, Leclerc, Geman-McClure... Nous avons opté pour la fonction de Tukey, présentée à la figure 3.6, car nous souhaitons que les outliers passant un certain seuil ne jouent absolument aucun rôle dans l'estimation des mouvements.

La comparaison de son tracé avec celui de la fonction quadratique met en évidence deux comportements diamétralement opposés vis à vis des outliers :

- dans le cas de la fonction de Tukey, les outliers ont un coût constant, et l'application de cet estimateur ne conduit pas à "distordre" le modèle pour essayer de les prendre en compte.

⁵C'est-à-dire s'il y a exactement deux mouvements transparents sur l'image.

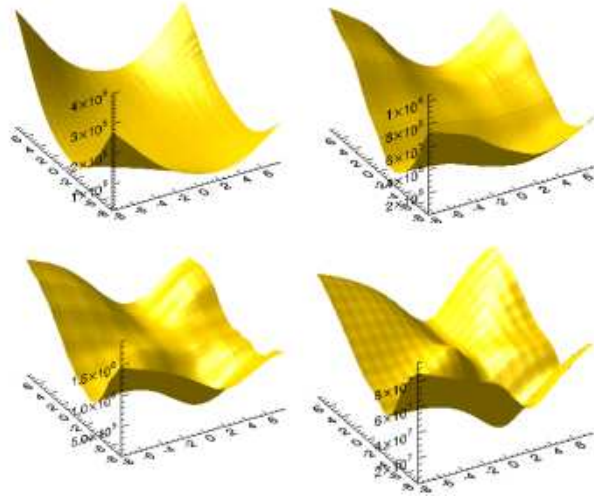


FIG. 3.5: Tracé de l'énergie J à différents niveaux de la pyramide de résolution, en fonction des paramètres de mouvement, pour une séquence d'images transparentes dont les couches sont en translation suivant l'axe des abscisses. L'espace de représentation est restreint aux deux composantes translationnelles $a_{1,1}$ et $a_{1,2}$, toutes les autres étant fixées à 0. De gauche à droite et de haut en bas, les niveaux 3, 2, 1 et 0 de la pyramide.

- à l'inverse, les outliers ont un impact important sur la fonction de coût quadratique, ce qui entraîne une "distorsion" du modèle estimé.

La fonction de Tukey est définie analytiquement comme suit :

$$\rho(r, C) = \begin{cases} \frac{r^6}{6} - \frac{C^2 r^4}{2} + \frac{C^4 r^2}{2} & \text{si } |r| < C \\ \frac{C^6}{6} & \text{sinon} \end{cases} \quad (3.29)$$

Elle dépend d'un paramètre d'échelle C , qui définit le seuil sur r pour lequel le point correspondant sera considéré comme un outlier. Puisque le contenu des séquences cliniques nous est inconnu, nous ne pouvons pas déterminer cette valeur a priori.

Nous avons choisi de déterminer C en ligne, de la manière suivante :

$$\begin{cases} \bar{r} &= \operatorname{med}_{\mathbf{p} \in \mathfrak{S}} r(\mathbf{p}, \theta_1, \theta_2) \\ \bar{\Delta}r &= 1.48 \times \operatorname{med}_{\mathbf{p} \in \mathfrak{S}} |r(\mathbf{p}, \theta_1, \theta_2) - \bar{r}| \\ C &= 2.795 \times \bar{\Delta}r \end{cases} \quad (3.30)$$

L'utilisation des médians permet d'obtenir une valeur typique \bar{r} de r puis de l'écart moyen $\bar{\Delta}r$ à \bar{r} , sur lesquels les outliers n'ont pas d'impact. La constante 1,48 permet de débiaiser l'estimateur de l'écart-type à partir de $\bar{\Delta}r$ et la constante 2,795 a été proposée par Tukey pour débiaiser l'estimation de C [Holland 77].

Ce calcul automatique de C a l'avantage majeur d'éviter d'avoir à régler de manière manuelle ce paramètre, d'autant plus qu'il peut varier notablement d'une séquence à l'autre (et même du début de l'estimation à la fin). Il a par contre l'inconvénient de nécessiter le calcul de deux opérateurs médians sur les pixels de l'image.

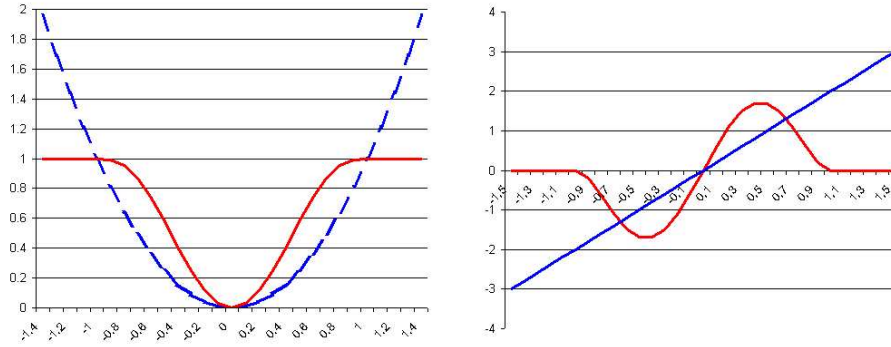


FIG. 3.6: Tracé de la fonction de Tukey (en rouge), comparé à celui de la fonction quadratique (en bleu). À gauche les fonctions en question, à droite leurs fonctions d'importance. Il s'agit de la dérivée de la fonction quadratique et de celle de Tukey.

Une implantation future en temps réel devra probablement opter pour une autre stratégie quant au choix de C . Nous pouvons imaginer un réglage depuis un tableau de valeurs de références (établies à partir de tests), dépendant :

- du bruit dans l'image (auquel nous avons accès par la mesure du rayonnement émis puis reçu),
- de l'angulation et de la modalité utilisées (donnant une indication sur l'anatomie imagée et donc les mouvements attendus),
- et du stade de l'estimation auquel nous nous trouvons (qui nous renseigne sur la fiabilité des champs de mouvement estimés jusque là).

3.3.3 Minimisation robuste par la méthode des moindres carrés pondérés itérés

Nous avons à minimiser la fonction :

$$J_r(\Theta) = \sum_{\mathbf{p} \in \mathfrak{S}} \rho(r(\mathbf{p}, \theta_1, \theta_2), C) \quad (3.31)$$

$$\begin{aligned} \text{avec } r(\mathbf{p}, \theta_1, \theta_2) = & I(\mathbf{p}, t+1) - I(\mathbf{p} + \mathbf{w}_{\theta_1}, t) \\ & - I(\mathbf{p} + \mathbf{w}_{\theta_2}, t) + I(\mathbf{p} + \mathbf{w}_{\theta_1} + \mathbf{w}_{\theta_2}, t-1) \end{aligned} \quad (3.32)$$

les composantes des vitesses u_{θ_1} , v_{θ_1} , u_{θ_2} et v_{θ_2} étant liées aux paramètres de mouvement par les deux relations (3.3).

Des problèmes voisins ont été résolus dans le domaine de l'estimation du mouvement, que ce soit pour l'estimation du champ dense des vitesses (non transparentes) [Black 93], pour l'estimation de modèles paramétriques de mouvement (non transparents) [Odobez 95], ou pour l'estimation de champs denses de mouvements transparents [Pingault 02b, Toro 00b]. [Holland 77] montre qu'il est équivalent de minimiser (3.31) et de considérer le problème

des moindres carrés pondérés itérés dont la formulation à l'itération j est la suivante :

$$\min_{\Theta^j} J_r^j(\Theta) = \sum_{\mathbf{p} \in \mathfrak{S}} w^j(\mathbf{p}) r(\mathbf{p}, \theta_1, \theta_2)^2, \quad (3.33)$$

$$\text{avec } w^j(\mathbf{p}) = \frac{\rho'(r(\mathbf{p}, \hat{\Theta}^{j-1}))}{2 \cdot r(\mathbf{p}, \hat{\Theta}^{j-1})}. \quad (3.34)$$

Chaque pixel se voit ainsi allouer un poids $w^j(\mathbf{p})$, d'autant plus grand qu'il se comporte en conformité avec le modèle de mouvement $\hat{\Theta}^{j-1}$ estimé à l'itération précédente. Il s'agit donc, en commençant avec des poids unité, alternativement

- de minimiser itérativement (3.33). Nous avons pour cela recours à l'une des méthodes de la section 3.2,
- de mettre à jour les poids avec (3.34).

En plus de résoudre le problème de l'estimation de manière robuste, cette approche fournit aussi une carte des poids qui met en évidence les outliers. Elle nous sera utile pour décider du nombre de couches dans notre méthode de segmentation développée au chapitre 5.

3.4 Initialisation

La méthode robuste d'estimation multirésolution précédente est rapide et efficace, mais elle est aussi sensible à l'initialisation. L'étape d'initialisation s'avère être cruciale pour la fiabilité de l'estimateur de mouvement. C'est ainsi que nous avons développés trois alternatives au cours de la thèse, pour aboutir sur une approche menant à des précisions tout à fait satisfaisantes. En outre, l'initialisation est d'autant plus importante ici qu'il convient de "désymétriser" le rôle des deux vecteurs de paramètres θ_1 et θ_2 . Nous avons en effet souligné que dans toutes les étapes présentées dans les deux sections précédentes, les rôles de θ_1 et de θ_2 étaient parfaitement symétriques. Si les champs de mouvement étaient initialisés de la même façon, ils seraient traités exactement de la même manière du début à la fin de l'algorithme et donc *resteraient égaux en fin d'estimation*. L'initialisation que nous proposons permet d'éviter cette situation.

Nous avons étudié différentes alternatives pour ce faire, recherchant des estimateurs rapides et peu sensibles au bruit. Nous pouvons par contre accepter des estimations imparfaites à ce stade de l'algorithme. Nous commençons en évoquant la méthode du simplex, que nous avons tout d'abord privilégiée. Nous avons ensuite constaté dans nos travaux qu'une approche en deux temps, mettant en jeu une mise en correspondance par blocs suivie d'un calcul des modèles affines qui leur correspondent, était plus efficace.

Seulement, cela suppose que les déplacements calculés par la mise en correspondance par blocs soient préalablement séparés en deux groupes (correspondant aux deux couches). Nous avons imaginé deux méthodes de séparation des champs. La première organise les vecteurs de déplacements calculés par la mise en correspondance par blocs en deux groupes, en utilisant une méthode de type $k - NN$ glissant, puis extrait de manière robuste les paramètres des modèles affines correspondants.

La seconde s'appuie sur une transformée de Hough dans un espace affine simplifié à trois paramètres. L'incrémentement de la matrice d'accumulation tient compte de la confiance associé aux vecteurs de la mise en correspondance par blocs. Nous détaillons ces trois approches dans les sous-sections suivantes.

3.4.1 Initialisation par la méthode du simplexe

Une approche intéressante est d'utiliser la méthode du *downhill simplex* [Press 92]. Il s'agit d'une méthode heuristique de minimisation de fonctions non linéaires, que nous pouvons appliquer à J pour initialiser notre algorithme. Elle consiste à déplacer itérativement $n + 1$ points (définissant le simplexe) dans l'espace de dimension n sur lequel est défini J . Dans notre cas, nous considérons donc 13 échantillons.

Il s'agit de remplacer itérativement le point du simplexe qui donne la valeur de J la plus élevée par un nouvel échantillon, recherché symétriquement à l'hyperplan défini par les n points conservés. Des mécanismes de dilatation (augmentant le volume du simplexe) ou de contraction (diminuant le volume du simplexe) sont proposés pour choisir la position du nouvel échantillon sur la droite ainsi définie. L'algorithme s'arrête quand le simplexe occupe un volume inférieur à un seuil donné.

L'idée est que, si la fonctionnelle à minimiser est raisonnablement régulière, cette stratégie de minimisation permet de visiter intelligemment l'espace de recherche. Dans nos contributions [Auvray 05a, Auvray 05b, Auvray 05c], nous l'avons mise en oeuvre sur le niveau supérieur de la pyramide de résolution pour des raisons de temps de calcul.

Nous avons depuis établi qu'une méthode en deux étapes, commençant par une mise en correspondance par blocs adaptée à la bi-transparence, et suivie d'une séparation des champs de déplacements, donne de meilleurs résultats [Auvray 06b].

3.4.2 Mise en correspondance par blocs pour la bi-transparence

Les méthodes d'initialisation les plus matures que nous avons développé s'appuient sur une approche en deux temps : calcul initial de vecteurs de déplacement par mise en correspondance par blocs (adaptée à la transparence), puis extraction des champs sous-jacents.

Nous calculons donc dans un premier temps un ensemble de vecteurs de déplacement pour une situation de transparence à deux couches (soit deux vecteurs par point) en reprenant l'approche de [Stuke 03a, Stuke 04]. Nous divisons l'image en blocs disjoints (pour des raisons de temps de calcul, et parce que nous ne nous intéressons qu'à une initialisation). Nous considérons que l'hypothèse de translation locale induite par un tel schéma reste raisonnable sur ces voisinages, et que les mouvements des deux couches peuvent alors être considérés comme translationnels sur chacun des blocs.

Pour déterminer les deux vecteurs de déplacement correspondant à un bloc donné, nous avons recours à une recherche exhaustive par discrétisation de l'espace de recherche. Plus précisément, nous calculons sur chaque bloc B_i

$$J_{B_i}(\mathbf{w}_1, \mathbf{w}_2) = \sum_{\mathbf{p} \in B_i} r(\mathbf{p}, \mathbf{w}_1, \mathbf{w}_2)^2 \quad (3.35)$$

$$\text{avec } r(\mathbf{p}, \mathbf{w}_1, \mathbf{w}_2) = I(\mathbf{p}, t + 1) - I(\mathbf{p} + \mathbf{w}_1, t) - I(\mathbf{p} + \mathbf{w}_2, t) + I(\mathbf{p} + \mathbf{w}_1 + \mathbf{w}_2, t - 1) \quad (3.36)$$

pour un ensemble de déplacements possibles $\mathbf{w}_1 \times \mathbf{w}_2$. Nous retenons les déplacements donnant l'énergie J_{B_i} la plus faible.

En pratique, nous avons établi par des observations de séquences cliniques que les déplacements des organes avaient une amplitude inférieure à 8 pixels. Notre espace de recherche

s'étend donc pour chaque coordonnée des vitesses de -8 pixels à 8 pixels, avec un pas de discrétisation de 1 pixel.

Les procédés de mise en correspondance par blocs fonctionnent en temps réel pour des images non transparentes. Seulement, nous devons dans le cas de la bi-transparence explorer un espace à 4 dimensions (une par composante des vecteurs de déplacement), contre 2 pour l'estimation de mouvements non transparents, ce qui rend le procédé beaucoup plus coûteux. Pour l'accélérer, nous avons décidé de l'appliquer à une pyramide de résolution (du type de celle introduite dans la sous-section 3.2.2).

Plus précisément, au niveau $h - 1$, \mathbf{w}_1^{h-1} balaye l'espace $[2.u_i^h - 2, 2.u_i^h + 2] \times [2.v_i^h - 2, 2.v_i^h + 2]$. Trois niveaux de résolution nous permettent d'explorer efficacement l'espace de recherche que nous nous sommes fixé. Après divers tests sur des images synthétiques de la section 1.3.2, nous avons établi qu'il était approprié de travailler sur des blocs de taille 32×32 pixels à *chaque niveau de la pyramide*. Aux niveaux les plus grossiers, nous n'appréhendons pas la composante de divergence des mouvements affines, mais une taille de bloc inférieure serait trop peu informative pour une initialisation fiable.

3.4.3 Séparation des champs de déplacement

Nous souhaitons maintenant extraire de cet ensemble les deux champs de mouvement sous-jacents. Pour l'exprimer de manière illustrée, il nous faut en quelque sorte passer de la figure de gauche de la Fig.3.7 à celle de droite. Nous proposons deux méthodes pour ce faire, selon que la séquence d'images traitée est très bruitée ou non.

La première d'entre elle s'appuie sur une estimation robuste des modèles de mouvement expliquant au mieux les vecteurs de déplacement observés. Seulement, il faut pour l'appliquer séparer au préalable l'ensemble indifférentié des vecteurs de déplacements en deux groupes, un pour chaque couche. Nous proposons pour ce faire un clustering de type k-NN glissant, mais nous observons qu'il conduit à des résultats peu fiables sur des données bruitées.

Nous proposons alors de manière alternative une extraction des modèles de mouvement par application d'une transformée de Hough. Nous nous intéressons à ce stade pour des raisons de stabilité à des modèles de mouvement affines simplifiés. Pour améliorer la robustesse de cette version de l'initialisation, nous avons recours à une incrémentation continue de la matrice d'accumulation, reflétant la confiance que nous accordons aux vecteurs de déplacements associés à chaque bloc.

3.4.3.1 Clustering et estimation robuste.

Estimation robuste des modèles de mouvement. La première approche consiste donc à calculer les deux modèles affines de mouvement correspondant au mouvement d'ensemble des deux couches transparentes à partir des déplacements calculés sur chacun des blocs. Nous pouvons considérer pour chacun d'eux un critère quadratique :

$$C(\theta_i) = \sum_{B_i \in B} (u_i - a_{1,i} - a_{2,i}x_{b_i} - a_{3,i}y_{b_i})^2 + (v_i - a_{4,i} - a_{5,i}x_{b_i} - a_{6,i}y_{b_i})^2 \quad (3.37)$$

où (u_i, v_i) est la composante du déplacement local correspondant à la couche i ($i = 1, 2$), et x_{b_i} et y_{b_i} sont les coordonnées du centre du bloc B_i ⁶. En fait, le calcul des déplacements

⁶Nous pouvons justifier le choix de ce critère de deux façons :

par mise en correspondance par bloc est imparfait, notamment sur les blocs peu texturés de l'image. Il conduit à un certain nombre de mesures inexactes qui risquent de fausser l'estimation des modèles affines. C'est pourquoi nous avons recours à un formalisme robuste et nous considérons finalement le critère suivant :

$$\min_{\theta_i} C_r(\theta_i) = \sum_{B_i \in B} \rho \left(\sqrt{(u_i - a_{1,i} - a_{2,i}x_{b_i} - a_{3,i}y_{b_i})^2 + (u_i - a_{4,i} - a_{5,i}x_{b_i} - a_{6,i}y_{b_i})^2} \right) \quad (3.38)$$

Cette minimisation est effectuée comme précédemment par les moindres carrés pondérés itérés. Il n'est pas nécessaire ici de déterminer le paramètre de Tukey en ligne puisque les grandeurs manipulées ont un sens physique aisément interprétable. Nous considérons que les vecteurs distants de plus de 2 pixels du champ estimé (au sens de $\|\mathbf{d} - \mathbf{w}_\theta\|$) sont des *outliers*, et fixons donc le paramètre de Tukey à 2.

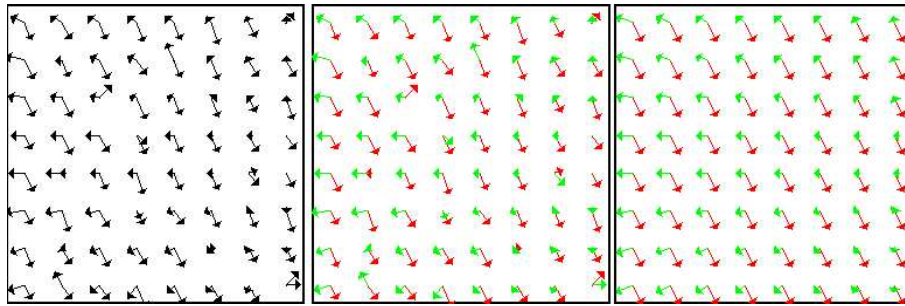


FIG. 3.7: De gauche à droite : le double champ de déplacement estimé par la mise en correspondance par bloc, les mêmes vecteurs attribués à chacun des deux champs, les modèles affines correspondants.

Clustering des vecteurs de déplacement. Seulement, cette approche suppose que l'ensemble des déplacements estimés sur chaque bloc ait été au préalable organisé en deux groupes distincts : celui qui correspond à la couche 1 et celui correspondant à la couche 2 (structure du centre de la Fig.3.7). S'il s'était agi de champs globalement translationnels, nous aurions pu avoir recours à un clustering de type $k - NN$ pour ce faire. Nous aurions itérativement calculé le barycentre de chacun des groupes, puis attribué les vecteurs sur chaque bloc de façon à minimiser la distance aux deux barycentres.

- on peut considérer que la mise en correspondance par blocs permet d'estimer le mouvement au centre du bloc. Dans cette optique, le calcul se fait sur un voisinage local autour de ce pixel seulement pour contourner le problème de l'ouverture. Il est alors logique de comparer les déplacements estimés avec les champs de mouvement sous-jacents *au centre de chaque bloc*,
- on peut également considérer que la mise en correspondance par blocs donne une estimée du déplacement moyen sur le bloc, qu'il convient de comparer au déplacement moyen induit par un champs de mouvement affine sur le bloc. Or, celui-ci est égal au déplacement généré par le modèle paramétrique au centre du bloc. On aboutit donc au même critère.

Puisque nous considérons des champs affines, nous proposons d'appliquer le clustering précédent de manière plus locale sur des fenêtres glissantes de l'image, typiquement formées de 3×3 ou 5×5 blocs. Sur chacune d'entre elles, nous considérons que les déplacements issus d'un même champ sont assez proches entre eux pour qu'il n'y ait pas de confusion possible avec le deuxième champ, et appliquons le clustering sur ces vecteurs. Cette étape préalable est représentée à la Fig 3.8, et l'ensemble de la méthode d'initialisation est résumée à la Tab.3.1.

Si le nombre de déplacements calculés erronés est trop important, cette méthode risque

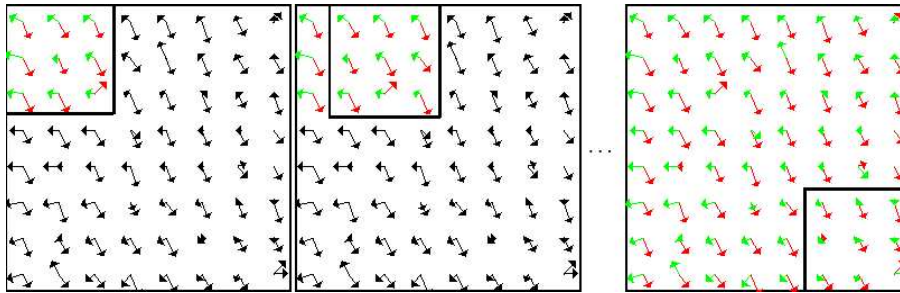
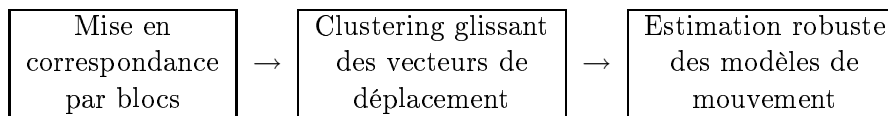


FIG. 3.8: De gauche à droite : les deux premières et la dernière itération du clustering glissant.

TAB. 3.1: Bloc-diagramme de la méthode d'initialisation par clustering glissant.



néanmoins d'échouer. La figure 3.9 illustre une situation où un déplacement calculé sur deux est faux. La clusterisation donnée par la technique glissante (structure au milieu de la Fig.3.9) est inexacte. Les vecteurs correspondant à la translation (champ rouge sur la structure de droite de la Fig.3.9) sont notamment attribués aux deux modèles sur la figure centrale. Cela risque de fausser l'initialisation.

Cette situation est typique de données très bruitées, comme dans les images fluoroscopiques. De nombreux blocs sont centrés sur des tissus peu contrastés, ce qui rend la mise en correspondance par blocs très sensible au bruit.

3.4.3.2 Transformée de Hough.

C'est pourquoi nous avons imaginé un deuxième mode d'estimation de champs affines qui permet de mieux initialiser notre méthode. Il cherche à s'appuyer essentiellement sur les blocs contenant des informations pertinentes.

Transformée de Hough sur un modèle simplifié. Nous proposons d'extraire les deux champs de mouvement présents dans cet ensemble de vecteurs translationnels en utilisant

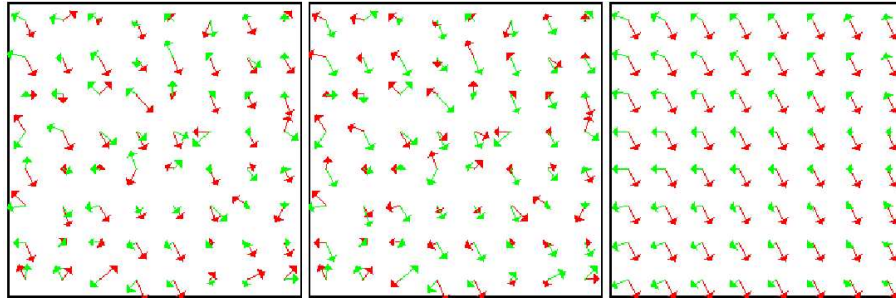


FIG. 3.9: Exemple de mauvaise clusterisation par la technique glissante, quand le nombre d'outliers est élevé. À gauche, les déplacements calculés correctement regroupés (vérité terrain). Au centre, le résultat du clustering glissant. Notons en particulier que les déplacements captant correctement la translation globale d'une des deux couches sont attribués aux deux modèles. À droite, les vrais champs affines.

la transformée de Hough [Illingworth 88]. Sur une image 288×288 (256×256 plus des marges de 16 pixels), en utilisant des blocs 32×32 , nous avons 128 contraintes à notre disposition pour extraire deux champs. Il n'est pas possible avec si peu de données de travailler de manière fiable dans l'espace de dimension 6 des paramètres de mouvement affine complet.

Nous avons donc décidé d'approcher les mouvements anatomiques de manière plus grossière dans le cadre de cette initialisation, en ne considérant que des modèles affines à trois paramètres : deux translations $a_{1,i}$ et $a_{4,i}$ et un facteur de divergence $a_{2,i} = a_{6,i} = h$. Cette approximation n'est pas sans fondement physique : les battements du coeur, la dilatation des poumons et les mouvements du diaphragme restent correctement pris en compte, en première approximation, par ce modèle plus simple.

Chaque vecteur de déplacement issu de la mise en correspondance par bloc doit donc vérifier :

$$\begin{cases} v_x = a_1 + hx \\ v_y = a_4 + hy \end{cases} \quad (3.39)$$

Cela se traduit par les contraintes suivantes sur les paramètres du modèle de mouvement :

$$\begin{cases} a_1 = hx - v_x \\ a_4 = hy - v_y \end{cases} \quad (3.40)$$

Ces équations définissent une droite dans l'espace (a_1, a_4, h) . Puisque deux champs de mouvement sont supposés être présents, l'ensemble des droites ainsi construit doit produire deux intersections nettes dans l'espace des trois paramètres de mouvement.

On considère en pratique la matrice d'accumulation correspondant à une discrétisation de cet espace à 3 dimensions et réceptionnant les votes satisfaisant les contraintes de (3.40). La matrice d'accumulation correspondant aux déplacements de la Fig.3.7 est illustrée à la Fig.3.10. Les deux intersections principales correspondent aux deux champs de mouvement en présence (marqués en vert et en bleu). La présence d'outliers dans les données ne perturbe pas cette procédure par transformée de Hough.

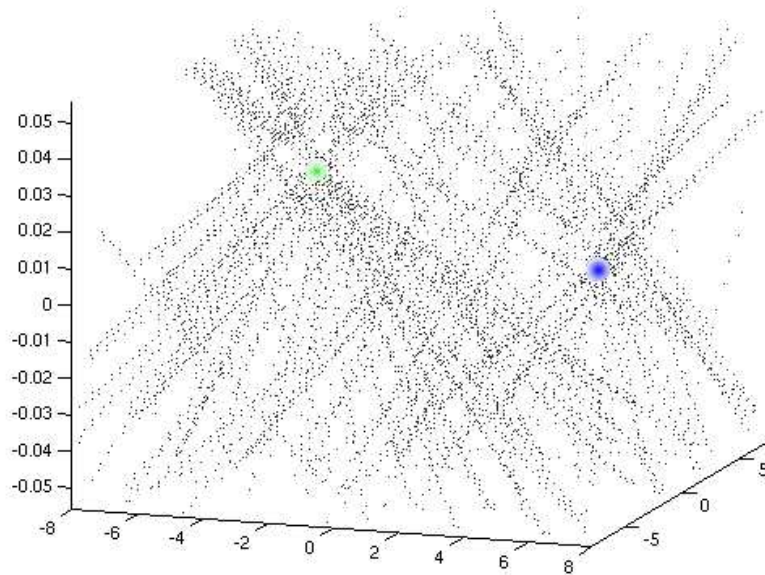


FIG. 3.10: Matrice d'accumulation dans l'espace (a_1, a_4, h) des paramètres de mouvement, construite à partir des vecteurs translationnels calculés sur l'image de gauche de la figure 3.7. La vérité-terrain des deux champs de mouvement présents y est représenté en vert et bleu.

Plus précisément, nous construisons une matrice d'*accumulation 3D*, correspondant à l'espace (a_1, a_4, h) , restreint aux intervalles $[-8, 8] \times [-8, 8] \times [-8/144, 8/144]$ sur les images de taille 288×288 ⁷ dans les exemples traités, et échantillonné au taux de 1 pixel pour les composantes translationnelles et de $1/144$ pour le facteur d'homothétie. Nous extrayons ensuite les deux champs de mouvement les plus représentés, en opérant d'abord au niveau de résolution inférieur de la matrice d'accumulation⁸ (pour détecter aussi des intersections significatives moins localisées).

Pondération des votes suivant leur pertinence. Afin de privilégier les mesures issues des blocs texturés, dont la fiabilité de calcul par la mise en correspondance par bloc est élevée, nous n'incrémentons pas la matrice d'accumulation de manière entière pour chaque cellule traversée par une droite. Nous affectons plutôt à chaque vote un coefficient continu dans $[0, 1]$ qui vaudra 1 si le bloc est suffisamment texturé, et 0 s'il ne l'est pas du tout. Nous ajoutons ce coefficient à chaque cellule traversée par la droite définie par les vecteurs de déplacement du bloc considéré.

⁷Puisque nous avons observé que l'amplitude des déplacements est inférieure à 8 pixels. Le facteur $1/144$ correspond d'autre part à la demi-largeur de l'image.

⁸La matrice d'accumulation au niveau de résolution inférieur a des dimensions moitié de celles de la matrice originale, et chacune de ses cellules est remplie par la somme des 8 cellules qui lui correspondent à pleine résolution.

Nous définissons ces coefficients comme suit. Considérons par exemple le vecteur \mathbf{w}_{θ_1} issu de la mise en correspondance sur le bloc B_i . Évaluons le résidu sur ce bloc :

$$J_{B_i}(\mathbf{w}_1, \mathbf{w}_2) = \sum_{\mathbf{p} \in B_i} r(x, y, \mathbf{w}_1, \mathbf{w}_2)^2 \quad (3.41)$$

$$\begin{aligned} \text{avec } r(\mathbf{p}, \mathbf{w}_1, \mathbf{w}_2) = & I(\mathbf{p}, t+1) - I(\mathbf{p} + \mathbf{w}_{\theta_1}, t) - I(\mathbf{p} + \mathbf{w}_{\theta_2}, t) \\ & + I(\mathbf{p} + \mathbf{w}_{\theta_1} + \mathbf{w}_{\theta_2}, t-1). \end{aligned} \quad (3.42)$$

r peut s'écrire, en notant I_1 la couche correspondant à \mathbf{w}_{θ_1} :

$$\begin{aligned} r(\mathbf{p}, \mathbf{w}_1, \mathbf{w}_2) = & I_1(\mathbf{p}, t+1) - I_1(\mathbf{p} + \mathbf{w}_{\theta_1}, t) \\ & - I_1(\mathbf{p} + \mathbf{w}_{\theta_2}, t) + I_1(\mathbf{p} + \mathbf{w}_{\theta_1} + \mathbf{w}_{\theta_2}, t-1) \\ & + I_2(\mathbf{p}, t+1) - I_2(\mathbf{p} + \mathbf{w}_{\theta_2}, t) \\ & - I_2(\mathbf{p} + \mathbf{w}_{\theta_1}, t) + I_2(\mathbf{p} + \mathbf{w}_{\theta_1} + \mathbf{w}_{\theta_2}, t-1). \end{aligned} \quad (3.43)$$

En supposant que \mathbf{w}_{θ_2} est correctement estimé, les deux dernières lignes de (3.43) disparaissent *quelles que soit la valeur de \mathbf{w}_{θ_1}* . Nous cherchons en fait à évaluer la pertinence du minimum de J associé à \mathbf{w}_1 .

Pour ce faire, nous comparons $J_{B_i}(\mathbf{w}_1, \mathbf{w}_2)$ à une valeur typique $J_{B_i}^{\bar{\mathbf{w}}_{\theta_1}}$ du critère J sur ce bloc, obtenue en moyennant J_{B_i} quand \mathbf{w}_{θ_1} varie (à \mathbf{w}_{θ_2} fixé).

- Si $J_{B_i}(\mathbf{w}_1, \mathbf{w}_2)$ et $J_{B_i}^{\bar{\mathbf{w}}_{\theta_1}}$ sont proches, on peut en conclure que l'impact de la couche 1 dans la minimisation de J est faible, ce qui révèle qu'elle est peu texturée, et donc que le calcul de \mathbf{w}_1 s'avère peu fiable.
- À l'inverse, si $J_{B_i}(\mathbf{w}_1, \mathbf{w}_2)$ est nettement inférieur à $J_{B_i}^{\bar{\mathbf{w}}_{\theta_1}}$, la couche 1 est certainement texturée et l'estimé \mathbf{w}_{θ_1} est fiable.

Plus formellement, nous calculons sur chaque bloc B_i :

$$c_1(B_i) = \frac{1}{N_B} \sum_{\Delta \mathbf{w}} J_{B_i}(\mathbf{w}_1 + \Delta \mathbf{w}, \mathbf{w}_2) - J_{B_i}(\mathbf{w}_1, \mathbf{w}_2), \quad (3.44)$$

$$c_2(B_i) = \frac{1}{N_B} \sum_{\Delta \mathbf{w}} J_{B_i}(\mathbf{w}_1, \mathbf{w}_2 + \Delta \mathbf{w}) - J_{B_i}(\mathbf{w}_1, \mathbf{w}_2). \quad (3.45)$$

$\Delta \mathbf{w}$ parcourant un jeu de valeurs prédéfinies (en pratique, $(-4, -4)$, $(-4, 4)$, $(4, -4)$, et $(4, 4)$)⁹.

Il reste à normaliser ces coefficients entre 0 et 1. Nous le faisons en calculant le premier quantile \tilde{c} des c_i , puis en affectant à chaque vote $c_i(B_i)/\tilde{c}$ (ou 1 si $c_i(B_i)/\tilde{c} > 1$). Ainsi, nous nous assurons que le quart des vecteurs les plus significatifs aura un coefficient valant 1, et que ceux qui sont moins informatifs, ou pour qui l'estimation n'est pas fiable, auront un coefficient faible. Ceci permet donc de se prévenir contre 75% d'outliers potentiels.

Nous montrons sur la figure 3.11 un exemple d'application de cette méthode sur un champ de mouvement affine simplifié bruité. Cette technique d'initialisation présente des coûts acceptables parce que le calcul des coefficients c_i se fait pendant la mise en correspondance par blocs, et parce que la transformée de Hough elle-même se fait au niveau des blocs (nombre restreint de mesures en jeu).

⁹L'utilisation de valeurs entières évite d'avoir à interpoler de nouvelles images !

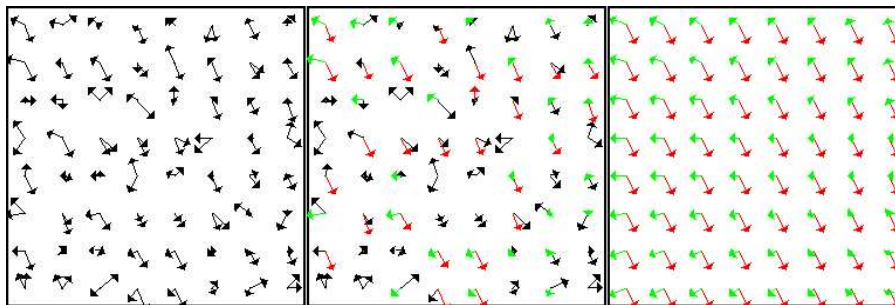


FIG. 3.11: Exemple d'estimation des deux modèles affines de mouvement (à 3 paramètres) par la transformée de Hough à partir d'un double champ de déplacements bruités. À gauche, les mesures considérées en entrée. Au centre, le résultat de l'extraction par transformée de Hough (sont laissés en noir les vecteurs non utilisés). À droite, la vérité-terrain pour cet exemple synthétique.

3.5 Résultats

L'algorithme 1 récapitule la méthode proposée pour estimer les mouvements transparents en situation de bi-transparence.

Nous présentons d'abord les résultats que nous avons obtenus sur les séquences synthétiques décrites à la section 1.3.2. Ils nous permettent d'évaluer objectivement la performance de la méthode définie et de justifier les réglages des paramètres de la méthode. Nous donnons par la suite des résultats sur des exemples cliniques.

3.5.1 Résultats synthétiques

Il nous faut évaluer les différentes options de réglage de notre estimateur de mouvements transparents. Même si en pratique les réglages optimaux ont été déterminés progressivement, nous choisissons dans cette section de ne pas respecter la chronologie des tests effectués. Nous préférons en effet présenter au lecteur des résultats isolant l'impact de chacun des facteurs autour des réglages finaux.

Les options de l'estimateur sont rassemblées avec leur réglage final dans le tableau 3.2.

Nos tests portent sur 250 séquences de trois images, générées comme expliqué en 1.3.2.

TAB. 3.2: Choix de réglage de l'algorithme

Paramètre	Réglage final
Encodage	Linéaire et racine carrée
Réduction de bruit	Filtre gaussien spatial 5×5 , $\sigma = 1.5$
Interpolation	Spline d'ordre 3
Initialisation	Transformée de Hough dans l'espace affine simplifié
Linéarisation	Second ordre

Algorithme 1 Algorithme d'estimation de mouvements transparents en situation de bi-transparence.

- **Initialisation**

- Calcul des deux déplacements sur chaque bloc par mise en correspondance par blocs (multirésolution).

- Extraction de champs de mouvement. Deux choix possibles.

1. Régression robuste après clustering.

- Clustering glissant de type $k - NN$,
- Extraction robuste des deux modèles par les moindres carrés pondérés.

2. Transformée de Hough.

- Calcul de coefficients de confiance sur chaque bloc,
- Remplissage de la matrice d'accumulation de Hough pour des modèles affines simplifiés à 3 paramètres (deux translations, un facteur d'homothétie),
- Extraction des paramètres des deux modèles affines (à 2 niveaux de résolution).

- **Estimation paramétrique robuste par la méthode de Gauss-Newton**

- Dans un cadre multirésolution, utilisation des moindres carrés pondérés pour minimiser itérativement une fonctionnelle, linéarisée autour de l'estimée courante. Deux possibilités :

1. Linéarisation au premier ordre.

- Minimisation par inversion matricielle.

2. Linéarisation au second ordre.

- Calcul d'une direction de descente par la méthode des gradients conjugués,
 - Calcul du pas par annulation d'un polynôme de degré 4.
-

Nous construisons l'exemple test avec vérité-terrain sur les mouvements transparents de la façon suivante. Nous formons une première couche en prenant une image issue d'un examen abdominal, que nous animons d'une translation. Nous y ajoutons une seconde couche avec une image d'un examen cardiaque, à laquelle nous appliquons un modèle affine simplifié. Pour déterminer le mouvement affine simplifié de façon réaliste, nous choisissons aléatoirement deux translations et un facteur d'homothétie h menant à un déplacement en chaque pixel compris entre -8 et 8 pixels. Nous étendons ensuite ce mouvement de référence à un mouvement affine complet en faisant varier chacun des paramètres $a_{2,k}$, $a_{3,k}$, $a_{5,k}$ et $a_{6,k}$ dans un intervalle de $20\%.h$ autour de leur valeur nominale ($a_{2,k} = a_{6,k} = h$, $a_{3,k} = a_{5,k} = 0$). Nous obtenons ainsi des mouvements caractéristiques des anatomies qui ne correspondent pas exactement au modèle pris en compte par la transformée de Hough. La nature des mouvements réels des organes étant inconnue, c'est à ce stade que nous faisons l'hypothèse la plus critique concernant le réalisme de nos séquences d'images tests. Néanmoins, les nombreux résultats convaincants sur séquences réelles laissent à penser que l'approximation "modèle affine complet" est correcte.

Nous imposons en outre aux deux champs de mouvements de différer d'au moins 2 pixels en moyenne sur l'image¹⁰. Nous interpolons les images avec des splines bicubiques.

Les images sont codées sur 12 bits, et leur valeur moyenne est typiquement de 500. Le traitement complet requiert une petite trentaine de secondes sur un Pentium IV (2.4 GHz and 1 Go) pour des images de taille 288×288 (256×256 , plus des marges de 16 pixels).

Nous synthétisons ainsi deux grands types de séquences :

- des images typiques de graphie (examen diagnostique), présentant un niveau de bruit d'écart-type $\sigma = 10$.
- des images typiques de fluoroscopie (examen interventionnel), présentant un niveau de bruit d'écart-type $\sigma = 20$.

Nous faisons également varier le taux de diffusé, qui peut être de 0%, 20% et 50%. Lorsque nous ne le précisons pas, il est fixé à la valeur typique de 20%. Nous ajoutons également un bruit électronique, d'écart-type $\sigma = 2$.

La métrique que nous présentons pour évaluer la performance de notre estimateur est *l'erreur moyenne d'estimation*, qui est la moyenne sur toute l'image de la différence entre le déplacement réellement appliqué et celui qui a été estimé, en pixels. Nous en donnons la valeur moyenne et l'écart-type sur l'ensemble des tests, et précisons, quand c'est pertinent, les temps de calcul.

Encodage. Nous commençons par comparer les performances de quatre types d'encodage.

Notre estimateur se base sur un modèle de formation de l'image qui suppose une transparence *additive*. Comme expliqué au paragraphe 1.2.2, une conversion non linéaire est indispensable pour passer d'images de rayonnement avec transparence multiplicative à des images avec transparence additive. L'opérateur naturel pour ce faire est le logarithme.

D'autre part, nous avons supposé implicitement que le bruit entachant nos images était gaussien, ce qui suppose l'utilisation sur les images originelles d'un opérateur racine carrée qui convertira le bruit quantique poissonnien en un bruit gaussien.

En outre, nous adjoignons à ces deux transformations possibles des variantes présentant

¹⁰Sans quoi un observateur ne percevrait qu'un seul mouvement dans la séquence en pratique.

une pente linéaire aux bas niveaux de gris (inférieurs à 224). Ceci permet une transition souple en 0 dans le cas du log, et évite de multiplier l'impact du bruit électronique aux bas niveaux de gris dans le cas de la racine carrée.

Les quatre transformations proposées sont présentées sur la Fig.3.12. Leur échelle a été choisie pour qu'elles aient des dynamiques voisines¹¹. De fait, les courbes tracées sont très proches pour la plage des niveaux de gris utiles ([300, 800]).

Les résultats des simulations sont regroupés dans le tableau 3.3. Ils montrent que la qualité des estimations est peu sensible au choix de l'opérateur d'encodage.

Nous optons finalement pour l'opérateur racine carrée avec partie linéaire, autant d'ailleurs parce que c'est l'opérateur traditionnellement utilisé au sein de General Electric que parce que les résultats sont légèrement meilleurs sur ces tests.

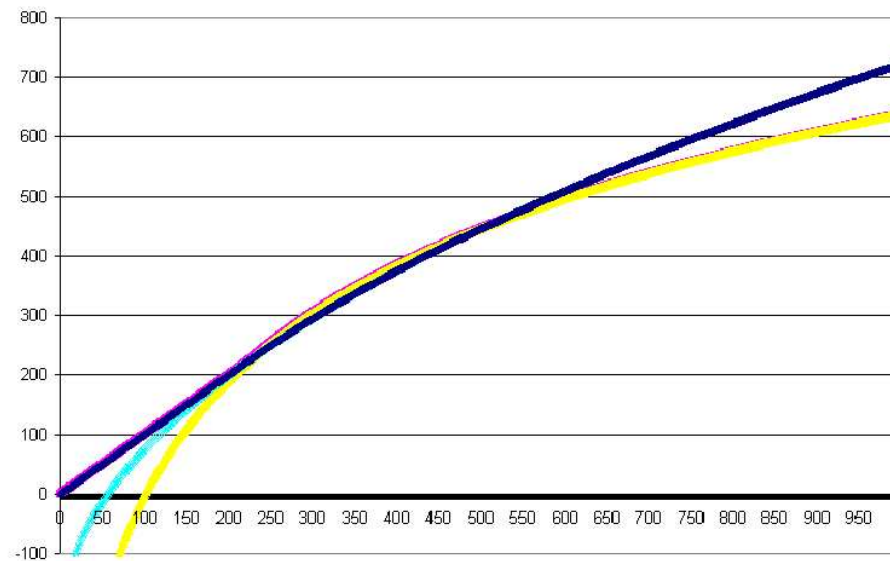


FIG. 3.12: Tracé des quatre opérateurs d'encodage testés. En bleu clair la racine carrée, en bleu foncé la racine carrée précédée d'une pente linéaire, en jaune le logarithme et en rose le logarithme précédé d'une partie linéaire.

Réduction de bruit. À l'origine, les images sont entachées d'un bruit qui est corrélé spatialement par la MTF du système d'imagerie. Rien ne dit que le filtrage spatial dû à la MTF (proche d'une gaussienne) offre un prétraitement optimal pour l'estimateur de mouvements transparents. Nous avons donc comparé la performance de notre estimateur sur les images brutes avec son comportement sur des images dont on a inversé la MTF, et auxquelles on a éventuellement appliqué un autre filtre spatial.

Ce prétraitement devant être rapide, nous nous sommes limités à l'utilisation de gaussiennes spatiales 5×5 de différentes largeurs. Nous craignons en effet que l'utilisation de

¹¹Plus précisément, elles transforment le niveau de gris X entrant 224 en 224, et 512 en 453. Ceci se traduit par une transformée logarithmique d'équation $Y = -1275 + 637,8 \log(X)$, et par une transformation racine carrée d'équation $Y = 2\sqrt{224X} - 224$.

TAB. 3.3: Tests des différents encodages. Nous rappelons que le niveau de diffusé est pour ces tests et les suivants fixé à 20%, et que le niveau de bruit est $\sigma = 10$ pour les séquences de graphie et $\sigma = 20$ pour les séquences de fluoroscopie.

Encodage	Erreur d'estimation de mouvement			
	Graphie		Fluoroscopie	
	Moyenne	Ecart-type	Moyenne	Ecart-type
Logarithme	0.27	0.66	0.95	1.21
Racine carrée	0.27	0.67	0.82	0.95
Logarithme avec partie linéaire	0.29	0.66	0.89	1.08
Racine carrée avec partie linéaire	0.27	0.66	0.82	0.94

filtres non linéaires (adaptatifs par exemple) induise des interactions complexes avec l'estimateur de mouvements transparents.

Dans la pratique, nous inversons la MTF en appliquant aux images avant encodage le filtre spatial inverse 9×9 (Fig.3.13). En tant que filtre de dimension finie, il ne permet pas une inversion parfaite de la MTF, mais nous avons vérifié qu'il mène à une décorrélation très satisfaisante du bruit, et qu'il avait des performances très proches d'une inversion dans le domaine de Fourier, tout en étant beaucoup plus rapide.

Les résultats des simulations sont donnés en Tab.3.4. Ils mettent en évidence que l'augmentation du niveau de bruit induit par l'inversion de la MTF n'est pas compensée par l'augmentation des contrastes apportée par ce passe-haut, ni par la décorrélation du bruit. Nous avons au contraire intérêt à lisser les images spatialement un peu plus fortement encore que la MTF ne le fait naturellement. Nous optons finalement pour un filtre spatial gaussien d'écart-type $\sigma = 1,5$.

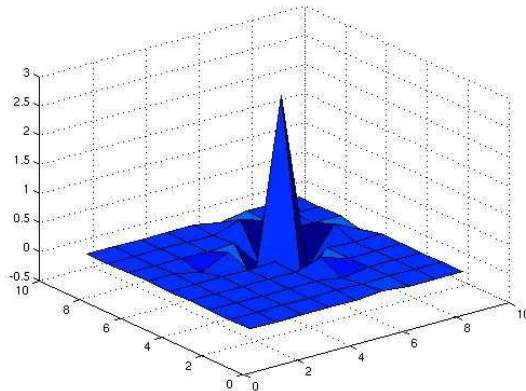


FIG. 3.13: Filtre 9×9 inverse de la MTF du système.

Interpolation. L'estimateur de mouvements bi-transparentes utilise de nombreux recallages d'images pour des déplacements non entiers lors de l'estimation paramétrique des

TAB. 3.4: Tests des différents prétraitements spatiaux

Prétraitement	Erreur d'estimation de mouvement			
	Graphie		Fluoroscopie	
	Moyenne	Ecart-type	Moyenne	Ecart-type
Pas de prétraitement	0.31	0.56	1.19	1.00
Inversion de la MTF	0.39	0.64	1.55	1.13
Inversion MTF + gaussienne $\sigma = 0.5$	0.38	0.62	1.47	1.05
Inversion MTF + gaussienne $\sigma = 1$	0.29	0.60	1.02	0.87
Inversion MTF + gaussienne $\sigma = 1.3$	0.24	0.60	0.90	1.06
Inversion MTF + gaussienne $\sigma = 1.5$	0.27	0.66	0.82	0.94
Inversion MTF + gaussienne $\sigma = 1.7$	0.28	0.67	0.96	1.20
Inversion MTF + gaussienne $\sigma = 2$	0.32	0.60	1.02	1.30

mouvements. Il est donc très dépendant du type d'interpolation choisi.

Le danger d'une interpolation grossière n'est pas tant qu'elle pose une limite intrinsèque à l'estimation, mais qu'elle peut introduire des minima locaux "parasites" sur les mouvements calculés et donc induire un biais. Un bruit non corrélé spatialement aura une amplitude plus faible pour les points interpolés de l'image, ce qui peut défavoriser des déplacements entiers. Ce phénomène est illustré à la Fig.3.14 pour le cas défavorable d'une translation (non entière). Dans le cas de mouvements affines, les positions visitées sont si nombreuses que cet effet est moyenné et donc beaucoup moins perceptible.

Ce phénomène est d'autant plus important que le niveau de bruit est élevé et le type d'interpolation grossier. Cependant, une interpolation élaborée est plus exigeante en temps de calcul.

Nous présentons les résultats obtenus en utilisant des splines d'ordre 2, 3 et 4 dans le tableau 3.5. Il n'est par contre pas possible de tester l'interpolation bilinéaire avec les mêmes réglages. Nous nous appuyons en effet sur une linéarisation au second ordre de (3.2), ce qui mène au calcul de dérivées spatiales d'ordre 2 des images. Or, une interpolation spline de degré n nous donne accès à une expression analytique des dérivées secondes sous forme d'un polynôme de degré $n - 2$. Dans le cas bilinéaire, les dérivées secondes analytiques sont nulles, et nous devons alors nous contenter d'une linéarisation au premier ordre de (3.2), ou bien calculer les dérivées secondes par différence finie. Nous avons fait cette dernière manipulation : elle mène à des précisions de 1,5 pixels sur des images fluoroscopiques, pour un temps de traitement de 20 secondes, et n'est donc pas compétitive avec les autres interpolations consignées dans Tab.3.5.

Nous pouvons noter que les précisions obtenues sur les images de graphie sont équivalentes, mais qu'il y a par contre un avantage perceptible à utiliser des splines d'ordre 3 sur les images fluoroscopiques, d'autant plus que le surcoût en temps de calcul est faible (+8%).

Initialisation. Nous testons ici les options relatives à une étape majeure dans l'algorithme d'estimation des mouvements transparents que nous proposons : le type d'initialisation. Nous comparons dans le tableau 3.6 le simplex, l'estimation robuste des paramètres

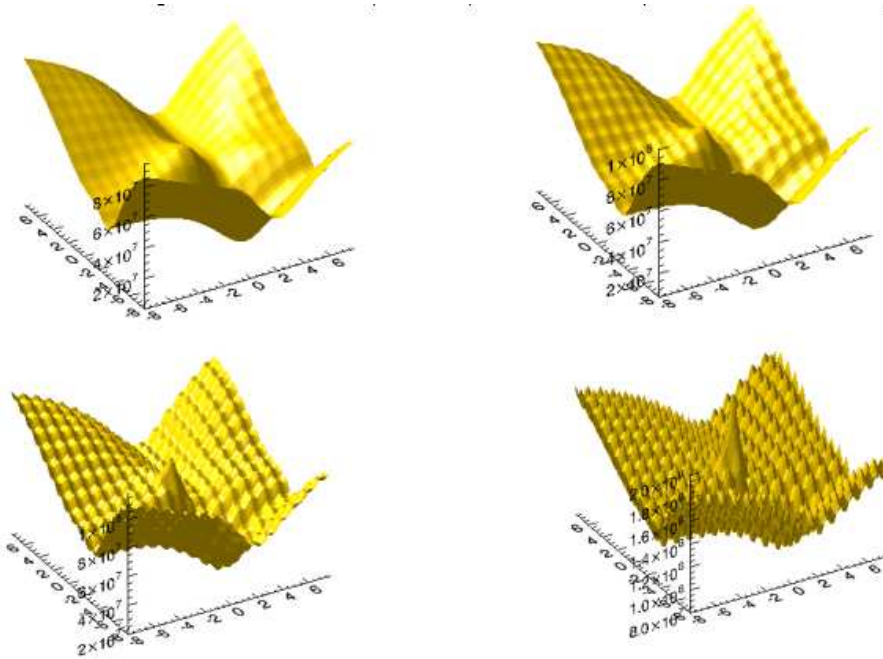


FIG. 3.14: Tracé de la fonctionnelle J en fonction des paramètres de mouvement pour une séquence d'images transparentes dont les couches sont en translation suivant l'axe des abscisses, et pour différents niveaux de bruit. L'espace de représentation est restreint aux deux composantes translationnelles $a_{1,1}$ et $a_{1,2}$, toutes les autres étant fixées à la valeur réelle du mouvement. De haut en bas et de gauche à droite, J pour une séquence non bruitée, pour un bruit d'écart-type $\sigma = 5$, $\sigma = 10$ et $\sigma = 20$. L'accroissement du pic en 0 s'explique par le fait que le bruit apporté par $I(\mathbf{p} + \mathbf{w}_{\theta_1}(\mathbf{p}), t)$ et $I(\mathbf{p} + \mathbf{w}_{\theta_2}(\mathbf{p}), t)$ n'y est pas indépendant et a donc une contribution deux fois plus forte qu'ailleurs.

du modèle de mouvement sur les déplacements issus de la mise en correspondance par blocs, préalablement organisés par un clustering glissant sur des fenêtres de 3×3 et 5×5 blocs, et une initialisation par transformée de Hough dans un espace affine simplifié.

La supériorité de l'approche s'appuyant sur la transformée de Hough ne fait pas de doute, alors même que les champs de mouvement considérés ne sont pas exactement des modèles affines simplifiés. Le clustering glissant se comporte mieux que le simplex, qui a par contre l'avantage d'être plus rapide (puisque'il n'intervient qu'au niveau le plus grossier de la pyramide de résolutions).

Linéarisation. Enfin, nous confrontons les résultats de l'estimation paramétrique multirésolution se basant sur une linéarisation au premier ordre de J à celle exploitant une linéarisation au second ordre de J . Comme évoqué plus haut, la première a l'avantage d'être rapide et sans minima locaux, et la deuxième de permettre d'atteindre des minima plus éloignés.

Nous avons réglé les algorithmes de manière très favorable pour le premier ordre, puisque nous n'avons pas pris garde au temps de calcul en autorisant 20 itérations du critère li-

TAB. 3.5: Tests des différentes interpolations

Spline	Erreur d'estimation de mouvement				Temps de calcul
	Graphie		Fluoroscopie		
	Moyenne	Ecart-type	Moyenne	Ecart-type	
Degré 2	0.28	0.63	0.92	0.96	26.8
Degré 3	0.27	0.66	0.82	0.94	28.9
Degré 4	0.31	0.76	0.81	1.08	30.8

TAB. 3.6: Tests des différentes initialisations

Initialisation	Erreur d'estimation de mouvement				Temps de calcul
	Graphie		Fluoroscopie		
	Moyenne	Ecart-type	Moyenne	Ecart-type	
Simplex	0.58	0.94	1.89	1.53	24.7
Clustering glissant 3×3	0.43	0.89	1.54	1.57	28.5
Clustering glissant 5×5	0.44	0.92	1.46	1.55	28.3
Transformée de Hough	0.27	0.66	0.82	0.94	28.9

néarisé à chaque niveau de résolution (contre 3 itérations pour le réglage au second ordre). Malgré cela, les résultats du tableau 3.7 montrent que la précision est meilleure en travaillant avec des linéarisations au second ordre.

Résultats définitifs. Au final, avec les options détaillées et retenues dans cette partie, le comportement de l'estimateur pour différents niveaux de bruit dans l'image et pour différents taux de diffusé est évalué en Tab.3.8, pour cinq métriques d'erreur : la moyenne, l'écart-type, le médian, le taux d'estimations significativement erronées et le pourcentage d'outliers. Les estimations significativement erronées désignent les estimations de modèles affines avec des erreurs supérieures à 2 pixels, et les outliers à 3 pixels.

Nous notons d'excellents résultats sur les séquences de graphie. Les estimés restent très fiables sur des images typiques de fluoroscopie, tant que le niveau de diffusé n'est pas trop important. Par contre, quand les deux difficultés du bruit et du diffusé s'ajoutent, les

TAB. 3.7: Tests des différents degrés de linéarisation

Linéarisation	Erreur d'estimation de mouvement				Temps de calcul
	Graphie		Fluoroscopie		
	Moyenne	Ecart-type	Moyenne	Ecart-type	
Premier ordre	0.25	0.49	1.03	1.02	87
Second ordre	0.27	0.66	0.82	0.94	27

TAB. 3.8: Métriques d’erreur d’estimation du schéma proposé, à différents niveaux de bruit et taux de diffusé : moyenne, écart-type, médiane, et outliers (estimés menant à des erreurs supérieures à 3 pixels).

Métrique	Bruit	Moyenne		
		Pas de diffusé	20% diffusé	50% diffusé
Moyenne	Graphie $\sigma = 10$	0.22	0.27	0.36
	Fluoroscopie $\sigma = 20$	0.53	0.82	1.67
Écart-type	Graphie $\sigma = 10$	0.63	0.66	0.70
	Fluoroscopie $\sigma = 20$	0.78	0.93	1.65
Médian	Graphie $\sigma = 10$	0.08	0.10	0.14
	Fluoroscopie $\sigma = 20$	0.25	0.41	1.00
Erronés (% <i>Erreur</i> > 2)	Graphie $\sigma = 10$	2%	2.4%	4.8%
	Fluoroscopie $\sigma = 20$	3.6%	9.2%	27.6%
Outliers (% <i>Erreur</i> > 3)	Graphie $\sigma = 10$	2.0%	2.4%	2.8%
	Fluoroscopie $\sigma = 20$	2.4%	4.8%	16.0%

résultats deviennent moins bons.

L’écart-type de l’erreur d’estimation, son médian et les taux d’estimations erronées et d’outliers, permettent de cerner plus complètement le comportement de l’estimateur. Dans la figure 3.15, nous avons tracé les erreurs commises à chacun des 250 tests effectués. Une très large majorité d’estimés sont regroupés en un nuage compact, et quelques rares outliers s’en écartent. Ceci plaide pour la fiabilité et la répétabilité de la méthode proposée.

Relâchement de la contrainte de la constance du mouvement sur trois dates successives.

Nous cherchons maintenant à examiner la sensibilité de l’estimateur de mouvements transparents à l’hypothèse la plus contraignante qui a été introduite, à savoir que les mouvements devaient être constants sur les intervalles de temps $[t - 1, t]$ et $[t, t + 1]$. Pour tester l’impact d’une violation relative de cette hypothèse, nous avons tiré les mouvements affines Θ_A des deux couches pour former la séquence synthétique aux dates $t - 1$ et t comme expliqué plus haut. Par contre, nous appliquons ensuite entre t et $t + 1$ un second champ de mouvement Θ_B , dérivé de Θ_A en autorisant chacun des coefficients à varier de 10, 20 ou 30% autour des valeurs prises entre $t - 1$ et t .

Nous mesurons cette fois-ci les erreurs moyennes sur l’image entre le champ estimé Θ et Θ_A d’une part, et Θ et Θ_B d’autre part. Nous en présentons la moyenne et l’écart type pour 250 simulations au tableau 3.9.

Nous constatons une dégradation *progressive* de la qualité des estimations des mouvements. Ce constat est très satisfaisant car il montre qu’il n’est finalement pas essentiel de respecter exactement l’hypothèse de constance du mouvement sur trois dates successives. Les précisions obtenues sur images fluoroscopiques restent tout à fait acceptables, même lorsque les deux mouvements successifs varient de 30% d’un instant au suivant.

Nous observons également une curiosité avec une précision plus grande pour des mouvements variant de 20% que pour ceux différents de 10% (pour des images fluoroscopiques)! Ceci nous rappelle que nos métriques sont très utiles pour comparer différentes versions de

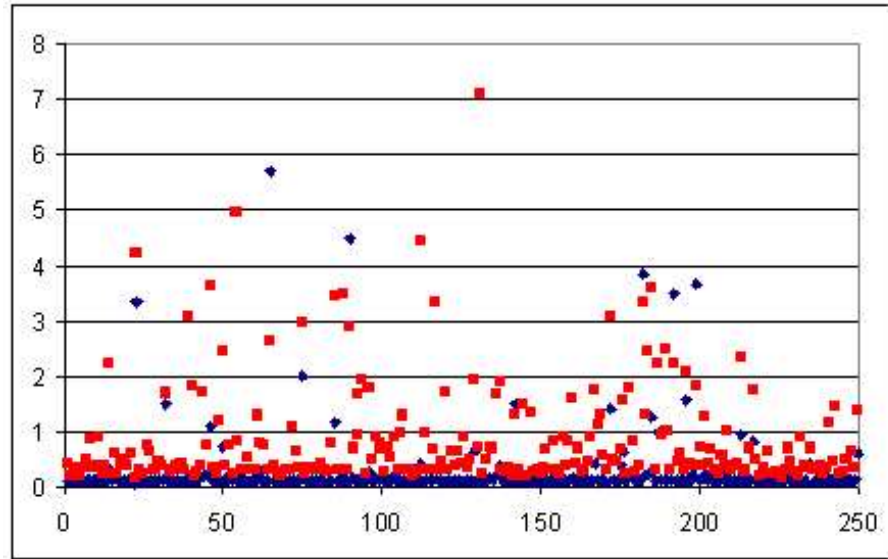


FIG. 3.15: Tracé des erreurs commises pour chacun des 250 tests définis précédemment, sur des images de graphie (en bleu, $\sigma = 10$) et de fluoroscopie (en rouge, $\sigma = 20$), avec 20% de diffusé.

TAB. 3.9: Moyenne et écart-type de l'erreur d'estimation pour différents niveaux de bruit et différentes variations des mouvements successifs (avec 20% de diffusé).

Métrique	Bruit	Variation des mouvements successifs			
		0%	10%	20%	30%
Moyenne	Graphie $\sigma = 10$	0.27	0.32	0.45	0.59
	Fluoroscopie $\sigma = 20$	0.81	1.00	0.92	1.07
Écart-type	Graphie $\sigma = 10$	0.66	0.49	0.58	0.45
	Fluoroscopie $\sigma = 20$	0.93	1.21	0.87	0.90

l'estimateur de mouvements transparents, mais sont également tributaires des tirages de modèles de mouvements effectués. Dans le cas qui nous intéresse, la médiane de l'erreur est bien plus faible pour des variations de 10% que de 20% (0.50 contre 0.60), mais quelques tirages "malheureux" ont occasionné la présence de 8.4% d'outliers¹² dans le premier cas contre 4.8% dans le second, ce qui fausse suffisamment la métrique de la moyenne pour que les fines différences de comportement mesurées soient gommées.

Notons également qu'étant donné les très bonnes précisions obtenues, notamment sur les images diagnostiques, la violation de l'hypothèse de translation locale est sans conséquence notable.

¹²C'est-à-dire menant à des erreurs d'estimation supérieures à 3 pixels.

3.5.2 Traitement de séquences d'images réelles

Nous avons extrait manuellement des portions de taille 288×288 de neuf séquences cliniques réelles (Fig.3.16), afin d'avoir des contenus de séquences d'images réelles en situation de bi-transparence sur lesquelles appliquer notre algorithme. Nous présentons nos résultats en traçant les champs de mouvements estimés¹³ et en affichant l'image correspondante de la séquence. Dans ces exemples réels, nous avons commencé par des situations vérifiant toutes les hypothèses d'utilisation de la méthode, pour aller progressivement vers des configurations plus difficiles.

Le lecteur peut se référer au site internet <http://www.irisa.fr/Vista/Equipe/People/Auvray/DemoBiTransparency.html> pour voir les vidéos concernées. Nous y présentons chaque séquence et montrons les champs de vitesse estimés à côté des séquences de trois images correspondantes pour quelques instants intéressants de la séquence. Nous donnons également les champs de mouvement estimés tout le long de la séquence afin de montrer la fiabilité de la méthode. Dans ce rapport, nous indiquerons pour chaque séquence traitée le taux approximatif d'estimation de mouvements cohérents avec l'observation.

3.5.2.1 Cas de condition normale de bi-transparence

La séquence fluoroscopique 17 (Fig.3.17) est typique d'une situation de bi-transparence : elle représente une partie du cœur, qui se superpose aux côtes. Les tissus des poumons sont également visibles, mais leur mouvement apparaît comme solidaire de celui du cœur sur la séquence extraite. Sous cette angulation, les mouvements de la paroi du cœur ressemblent à des translations souples, que le lecteur peut s'imaginer en suivant les tâches claires des tissus du poumon.

L'algorithme d'estimation des mouvements transparents donne des résultats très satisfaisants. Le mouvement du fond est bien évalué comme statique (le champ lui correspondant, presque nul, n'est parfois pas tracé), et le mouvement estimé des parois du cœur correspond à l'observation. Sur l'ensemble de la séquence, les vitesses estimées sont cohérentes avec l'observation pour environ 80% des images traitées.

Le séquence fluoroscopique 19 (Fig.3.18) présente également une configuration de bi-transparence nette. Nous y voyons le cœur (masse grise occupant les trois quarts de l'image, quasiment statique sous cette angulation) qui se superpose aux poumons (tissus clairs en haut à gauche). Le diaphragme (très sombre) est également présent dans la partie inférieure de l'image. La première couche, quasiment statique, correspondrait à l'ensemble cœur et côtes, et la seconde au groupe poumon-diaphragme. Soulignons ici l'intérêt d'avoir différencié la notion de couches de celle d'organes, qui conduit à une segmentation beaucoup plus pertinente au sens du mouvement.

L'originalité de cette séquence par rapport à la précédente tient au fait que le médecin déplace le système d'imagerie en cours d'acquisition pour le centrer sur les organes qui l'intéressent. Ce geste s'appelle le *panning*. Ce faisant, l'intégralité de l'image subit une translation globale.

Nous présentons à la Fig.3.18 une première image avant le panning, qui représente correctement les mouvements des deux couches. Les trois autres images présentées se réfèrent à

¹³Ces champs de mouvements sont orientés vers le passé. Ils sont multipliés par un facteur deux pour être plus facilement visibles.

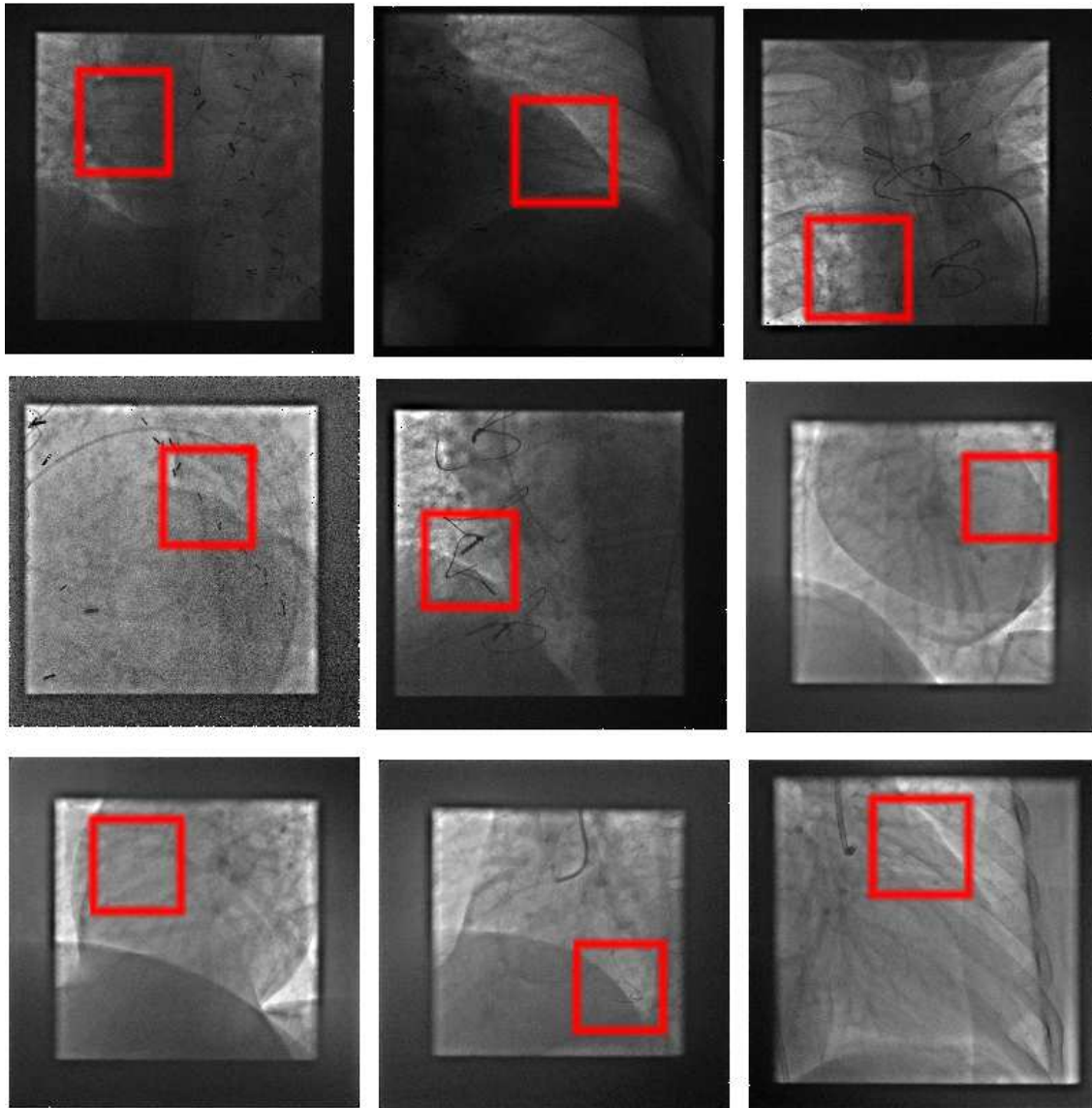


FIG. 3.16: Les neuf séquences dont sont extraites les portions traitées par la suite : séquences fluoroscopiques 17, 19, 3, 1, 6, séquences diagnostiques 301, 312, 33 et 405. Les localisations des imagettes 288×288 extraites sont indiquées par la fenêtre tracé en rouge.

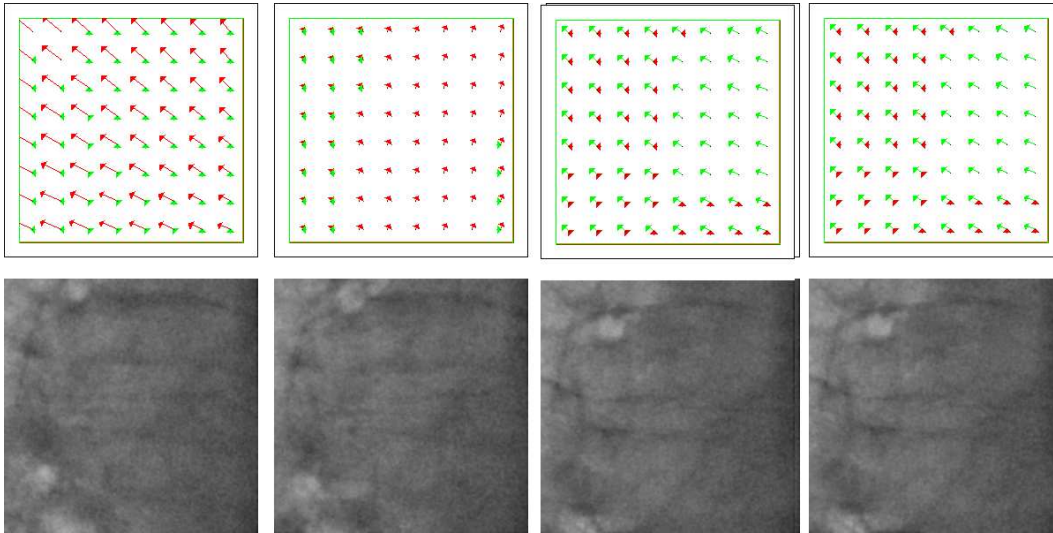


FIG. 3.17: Quatre champs de mouvements bi-couches estimés et les images correspondantes de la portion de la séquence fluoroscopique 17, aux dates 10, 20, 30, 33.

trois estimations successives lors du panning. Nous observons que le mouvement global de translation est bien appréhendé, et que les mouvements apparents des deux couches (statique entraîné par la translation globale, et “cœur et poumons” animés de leur mouvement propre et de la translation globale) restent cohérents avec l’observation. Cet exemple met aussi en évidence le fait que l’estimateur est capable d’estimer des mouvements d’amplitude relativement importante.

Le taux de bonnes estimation sur la séquence est de l’ordre de 70%. Les mauvaises estimations correspondent soit à des instants où les deux couches sont fixes (et alors une des deux vitesses estimées est nulle, mais l’autre est fantaisiste), soit à des moments du cycle cardiaque où l’hypothèse de constance temporelle du mouvement est niée.

La séquence fluoroscopique 3 (Fig.3.19) se situe elle aussi à la frontière entre le cœur et les poumons, sur un fond statique. Le cœur correspond à la masse sombre à droite de l’image, et les poumons aux tissus clairs à gauche. Ils sont animés d’un même mouvement en cet endroit de l’anatomie. On observe en outre une agrafe très contrastée à droite de l’image (provenant d’une intervention chirurgicale antérieure), associée au fond.

Cette séquence a la particularité d’imager le cœur dans une angulation sous laquelle ses mouvements rapides apparaissent clairement. La figure 3.19 montre que ses mouvements sont bien appréhendés que ce soit en phase de diastole (dilatation du cœur, à la date 21 par exemple) ou en phase de systole (contraction du cœur, à la date 34 par exemple). Le mouvement estimé pour le fond statique est bien quasi-nul.

Les estimations de mouvement sont correctes pour 80% de la séquence environ.

Enfin, la séquence fluoroscopique 1 (Fig.3.20) présente une configuration de bi-transparence impliquant un fond statique et un ensemble solidaire cœur-poumons. De petites agrafes sont visibles et sont solidaires du mouvement des poumons.

Les mouvements anatomiques sont bien estimés, mais cet exemple vaut surtout par le résultat de l’estimation à la date 41. Il met en effet en évidence que la méthode est tri-

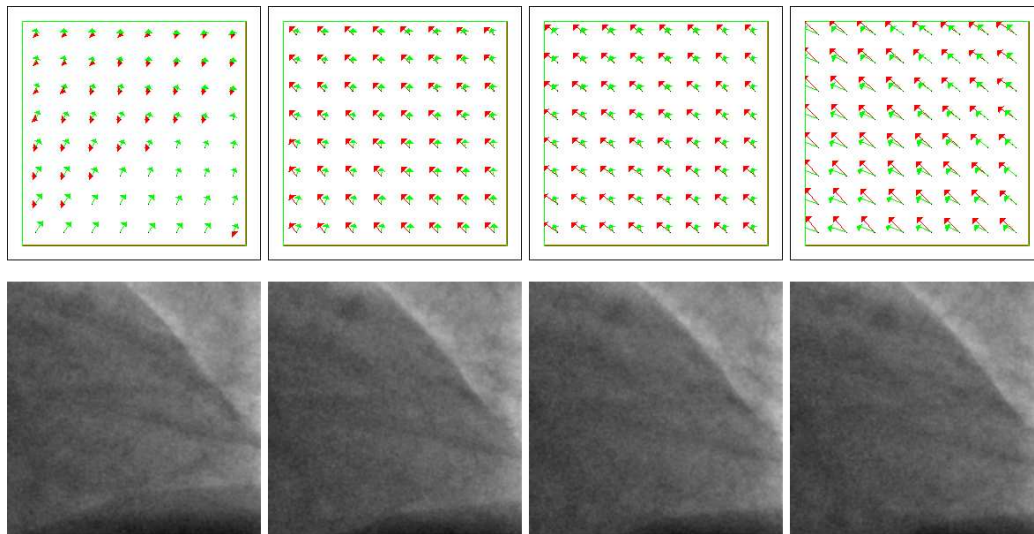


FIG. 3.18: Quatre champs de mouvements bi-couches estimés et les images correspondantes de la portion de la séquence fluoroscopique 19 – 1, aux dates 26, 38, 39, 40.

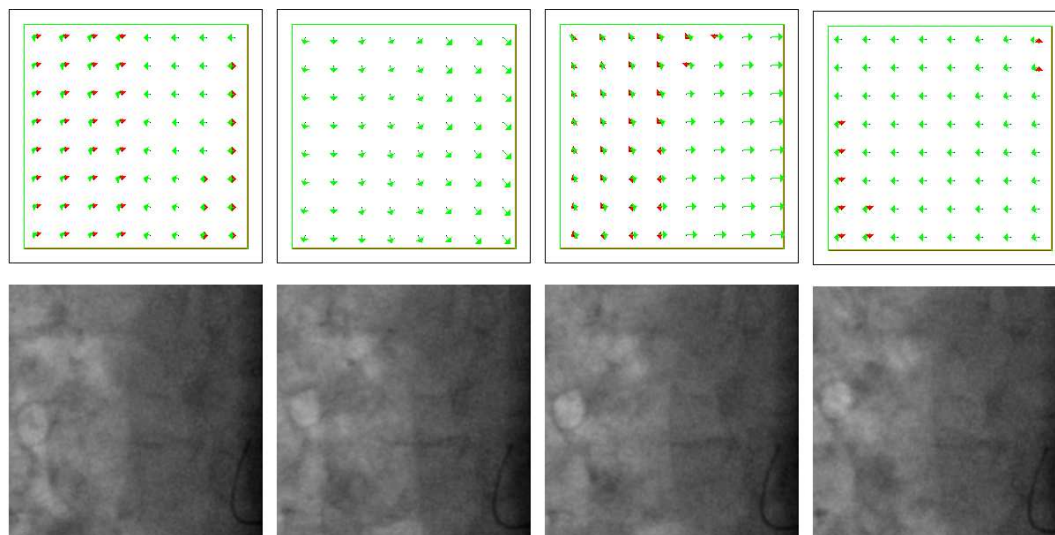


FIG. 3.19: Quatre champs de mouvements bi-couches estimés et les images correspondantes de la portion de la séquence fluoroscopique 3, aux dates 4, 17, 21, 34.

butaire d'un problème d'observabilité globale, classique dans toutes les problématiques d'estimation du mouvement. Nous sommes en mesure d'estimer avec précision le mouvement parallèlement aux gradients de l'image, mais nous n'avons aucun moyen localement de l'appréhender perpendiculairement à cette direction.

Comme nous considérons l'image entière, nous avons en général une distribution des directions de gradient suffisamment variée. Par contre, dans l'exemple présenté, la ligne de fort contraste du cœur permet d'accéder à l'information de mouvement selon une seule

direction de gradient de l'image. Les textures trop faibles contenues dans les autres zones de l'image ne donnent pas assez d'informations complémentaires.

Ainsi, si le mouvement de la couche mobile est bien estimé à la date 20, il est erroné à la date 41. Sur l'ensemble de la séquence, il est exact dans 80% des cas environ.

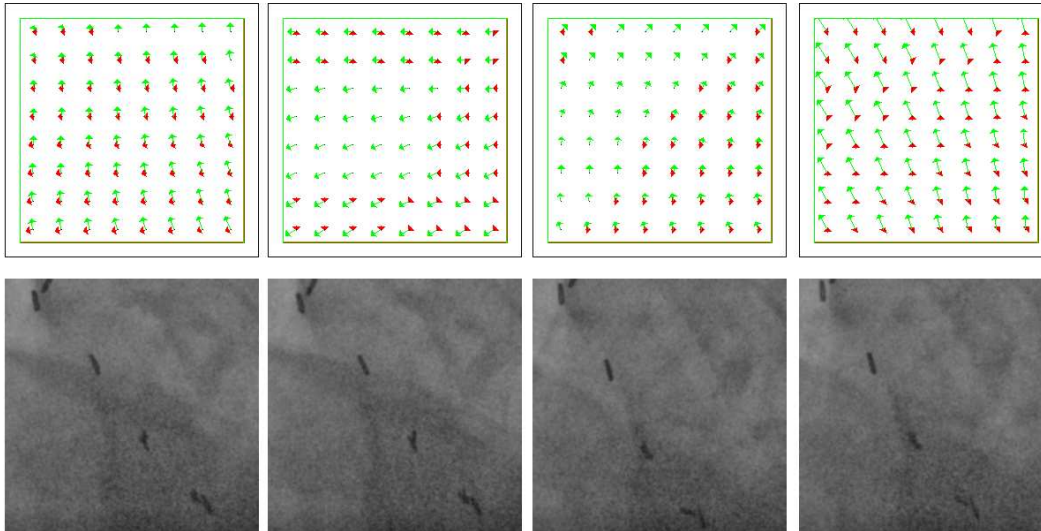


FIG. 3.20: Quatre champs de mouvements bi-couches estimés et les images correspondantes de la portion de la séquence fluoroscopique 1, aux dates 4, 7, 20, 41.

3.5.2.2 Cas aux limites de validité des hypothèses

Maintenant que la fiabilité de l'estimateur de mouvement en bi-transparence a été illustrée sur des séquences fluoroscopiques réelles, nous analysons son comportement sur des cas plus difficiles. Nous nous référons aux cas où certaines des hypothèses que nous avons faites ne sont plus valides.

Le premier exemple (séquence fluoroscopique 6 sur Fig.3.21) illustre le fonctionnement de l'algorithme en situation de *tri-transparence*. On aura d'ailleurs noté précédemment que les interprétations des images comme étant en situation de transparence bi-couche sont parfois sujettes à discussion. Dans quelle mesure les poumons ont-ils le même mouvement que le cœur? Le diaphragme que les poumons? La relative flexibilité du modèle affine permet de fusionner des couches ayant des mouvements proches en une même couche, mais qu'en est-il quand ils sont nettement différents?

Nous présentons sur la Fig.3.21 une séquence correspondant à une situation de transparence à trois couches :

- le fond statique, matérialisé notamment par les grosses agrafes visibles sur l'image,
- l'ensemble diaphragme-poumons. Le diaphragme est la masse très sombre occupant la moitié inférieure gauche de l'image, et les poumons la texture claire sur l'autre moitié de l'image. Du fait de leur proximité, leurs mouvements sont proches, tant et si bien qu'ils peuvent être considérés comme formant une seule couche.

- le cœur est également présent. Il s’agit de la masse gris clair convexe occupant la moitié droite de l’image. Il se distingue bien sur la quatrième image affichée.

Ce sont les mouvements des deux couches les plus contrastées qui sont estimés. Il s’agit dans ce cas du fond et de l’ensemble diaphragme-poumons aux dates 14 et 21. Cela s’explique par la nature robuste de l’estimateur de mouvements en bi-transparence, qui est capable d’ignorer les informations de mouvement venant du cœur pour retenir que les deux mouvements dominants.

Nous observons en outre qu’aux deux premières dates présentées (6 et 11), le diaphragme semble statique, ce qui est typique de son mouvement lent et ample. La couche lui correspondant est alors agrégée à celle du fond, et le mouvement du cœur est alors pris en compte.

Si l’estimation de mouvements en bi-transparence peut appréhender raisonnablement des situations réelles de trois couches en présence, l’une d’entre elles restant peu visible, il est évident que cela ne sera plus le cas si les trois couches sont également contrastées. Les estimés sont satisfaisants sur 80% de la séquence environ.

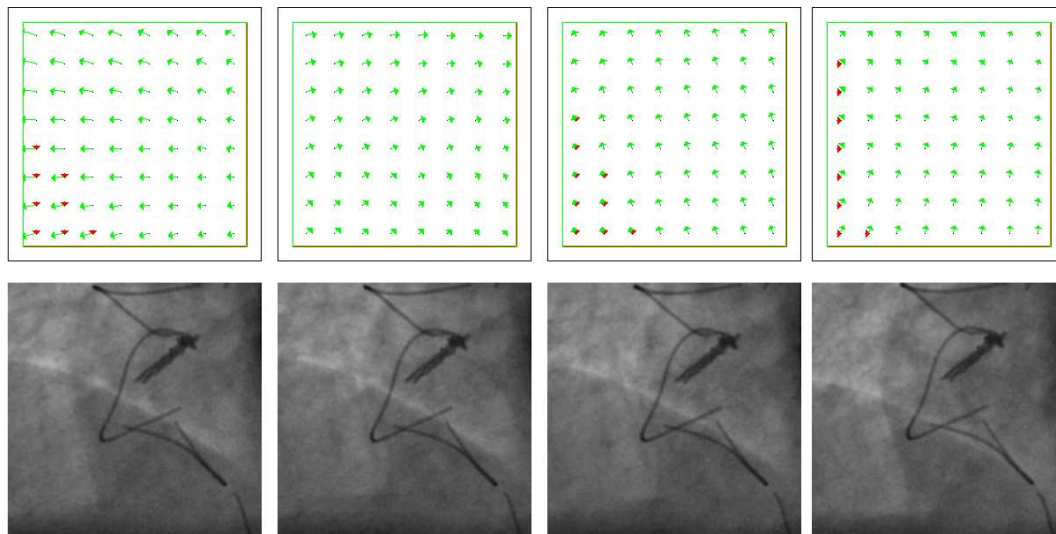


FIG. 3.21: Quatre champs de mouvements bi-couches estimés et les images correspondantes de la portion de la séquence fluoroscopique 6, aux dates 6, 11, 14, 21.

Sur la Fig.3.22, nous nous intéressons au comportement de l’algorithme lorsque le mouvement n’est pas constant sur deux intervalles de temps consécutifs. Nous considérons une séquence diagnostique en situation de bi-transparence : une couche correspond au cœur, et une autre aux poumons. Nous présentons quatre images successives, correspondant à des dates situées à la transition entre diastole et systole.

Si la première et la quatrième estimation du mouvement du cœur sont satisfaisantes, les deux autres estimations sont “fantaisistes”. C’est tout à fait compréhensible dans la mesure où l’hypothèse de constance temporelle du mouvement est alors complètement invalidée. Les deux mouvements successifs n’ont aucun rapport entre eux.

Remarquons par contre que le deuxième mouvement estimé reste correct. Cette séquence

est d'autant plus difficile que les tissus y sont particulièrement lisses, ce qui se traduit par un taux de bonne estimation de l'ordre de 50% environ sur l'ensemble de la séquence. Dans une optique applicative, il faudra détecter les ruptures de mouvements soit par des moyens externes, soit par des moyens internes à l'estimation de mouvement. Nous n'avons pas eu le temps d'aborder ce point qui sera évoqué dans les perspectives de ce travail.

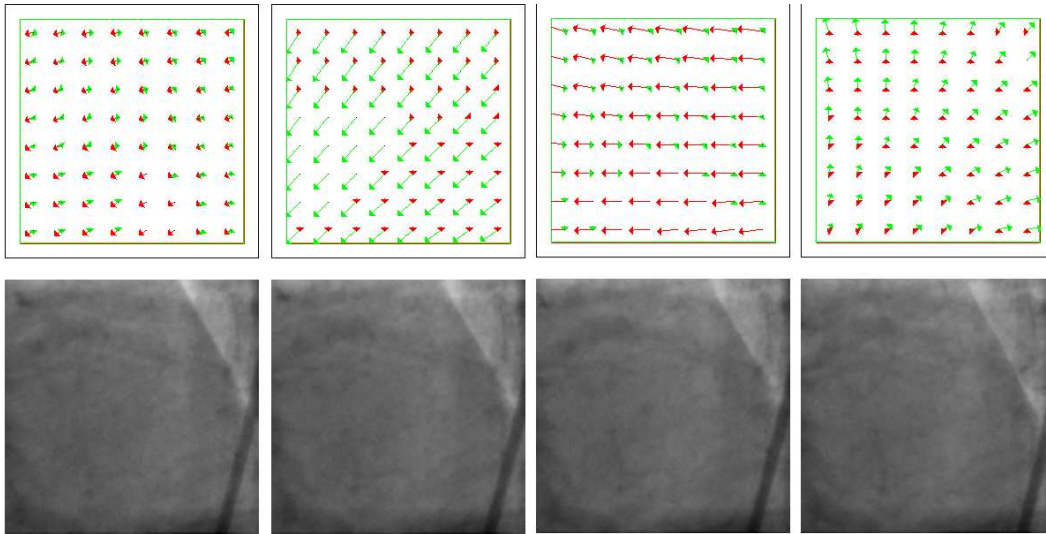


FIG. 3.22: Quatre champs de mouvements bi-couches estimés et les images correspondantes de la portion de la séquence diagnostique 301, aux dates 35, 36, 37, 38 (à cheval sur un cycle de diastole et de systole).

Nous étudions maintenant le comportement de l'estimateur de mouvements transparents lorsque des outils interventionnels sont présents à l'image, ou qu'une injection a lieu. Dans ce dernier cas, le médecin utilise le cathéter pour injecter un produit de contraste dans le réseau artériel du patient, afin de le rendre visible.

Le séquence diagnostique 312 est en situation bi-couche "cœur et poumons". Dans la première image, un cathéter a été inséré dans l'artère coronaire. Son mouvement est associé à celui du cœur. L'image suivante visualise les coronaires en cours d'injection. Ce phénomène ne perturbe pas l'estimation de mouvements transparents, en raison des estimateurs robustes utilisés.

Quand l'injection est stabilisée, elle contribue à enrichir les informations révélant le mouvement du cœur. La zone de perturbation tourbillonnaire au niveau du cathéter est ignorée, toujours grâce au critère d'estimation robuste.

Cette séquence reste difficile, surtout du fait des mouvements brusques des artères coronaires mises en évidence par le contraste, ce qui se traduit par des estimations fiables dans 60% des cas seulement.

Enfin, nous présentons sur la Fig.3.24 deux séquences diagnostiques illustrant la capacité de notre estimateur à appréhender les mouvements des outils interventionnels, grâce à l'estimation du mouvement des organes avec lesquels ils se déplacent. Il s'agit d'un guide

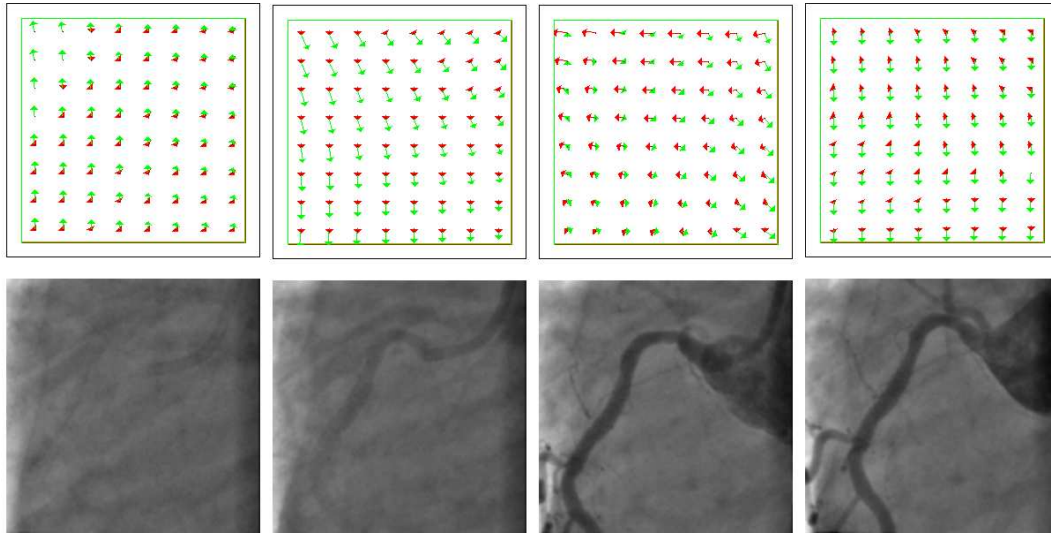


FIG. 3.23: Quatre champs de mouvements bi-couches estimés et les images correspondantes de la portion de la séquence diagnostique 312, aux dates 8, 26, 38, 42.

(séquence 33), et d'un stent (séquence 405), mis en évidence sur la Fig.3.24 par un cercle rouge.

Pour la première séquence traitée, les estimées sont cohérentes avec l'observation sur l'intégralité des images. La séquence du stent est plus difficile car les seuls mouvements en présence sont celui du fond d'une part, et celui du stent d'autre part. Malgré le fait qu'il occupe seulement une petite surface de l'image, il donne lieu à des estimées intéressantes sur 60% des instants traités.

3.6 Conclusion

Nous nous sommes intéressés dans ce chapitre à l'estimation de mouvement sur des séquences en situation de bi-transparence, c'est-à-dire comportant deux couches sur toute l'image. Ces couches correspondant à des organes (ou parties d'organes), elles sont animées de mouvements relativement simples, que nous modélisons par des modèles affines. Ainsi, nous pouvons contraindre adéquatement le problème posé et nous avons pu concevoir un estimateur fiable sur les images difficiles que nous considérons.

L'estimation des champs paramétriques est effectuée par une technique incrémentale de type Gauss-Newton. Cela permet de constamment se ramener au cas des petits mouvements, ce qui autorise à linéariser autour de l'estimé courant la fonction à minimiser. Deux options ont été étudiées : inversion matricielle en cas de linéarisation au premier ordre, ou méthode des gradients conjugués en cas de linéarisation au second ordre.

Pour que les estimations des mouvements transparents ne soient pas biaisées par d'éventuels outliers (il peut s'agir des outils interventionnels ou de petits mouvements parasites), nous avons recours à un formalisme robuste. Le problème se résout alors par la méthode des moindres carrés pondérés itérés.

Nous avons également étudié avec attention le problème de l'initialisation de ce schéma.

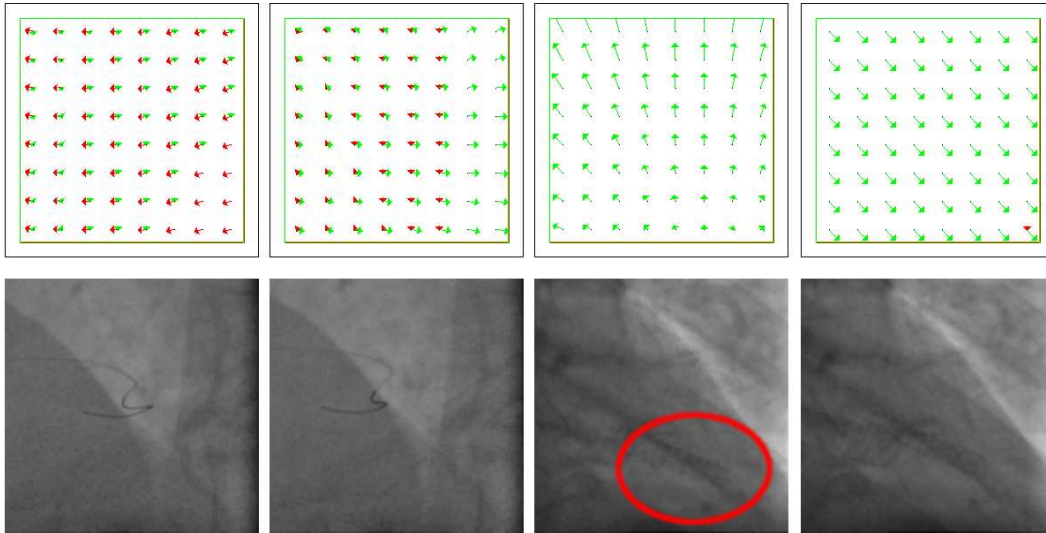


FIG. 3.24: Quatre champs de mouvements bi-couches estimés et les images correspondantes de la portion de la séquence diagnostique 33, aux dates 10 et 15, et de la portion de séquence diagnostique 405 aux dates 10 et 17.

Nous utilisons d'abord une mise en correspondance par blocs adaptée à la bi-transparence. Nous proposons alors deux méthodes pour extraire des deux ensembles de vecteurs translationnels résultants les deux champs affines de mouvement. La première consiste dans un premier temps à séparer les vecteurs de déplacement calculés en deux groupes correspondant aux deux couches par un *clustering* glissant de type $k - NN$, puis à calculer de manière robuste le champ de mouvement paramétrique correspondant à chacun des deux groupes.

Cette phase de *clustering* étant toutefois peu robuste aux erreurs de calcul de la mise en correspondance par blocs, nous proposons une alternative qui consiste à s'appuyer sur la transformée de Hough pour extraire de l'ensemble non structuré des vecteurs translationnels, les deux modèles affines de mouvement (dans une version simplifiée à ce stade). Pour des raisons de stabilité numérique, nous nous sommes en effet limités à un espace de Hough à 3 dimensions, correspondant à un modèle affine incluant deux composantes de translations et un facteur d'homothétie. Nous considérons que cette classe de transformations correspond grossièrement aux mouvements anatomiques, ce qui nous suffit dans le cadre de l'initialisation.

Nous avons alors mené de nombreux tests pour aboutir à un réglage optimal du schéma complet ainsi construit. Les images sont d'abord traitées par un filtre qui inverse la MTF, puis filtrées par une gaussienne spatiale 5×5 d'écart type $\sigma = 1.5$. Elles sont ensuite encodées de façon à obtenir une transparence additive par une transformation racine carrée (tronquée par une partie linéaire). La version finalisée de l'estimateur de mouvements transparents comprend une initialisation par transformée de Hough et une linéarisation de la fonctionnelle au second ordre d'estimation paramétrique. Les interpolations d'image utilisées font appel aux splines bicubiques.

Finalement, les résultats obtenus, que ce soit sur des séquences synthétiques réalistes ou sur

des images réelles, sont très satisfaisants, et assez peu sensibles à l'hypothèse de constance du mouvement sur deux intervalles de temps successifs, sauf en situation de totale invalidation de cette hypothèse.

La méthode définie est en outre capable de s'adapter à des situations de tri-transparence avec une couche peu contrastée, de suivre les outils interventionnels solidaires des organes, et d'être robuste à la réalisation d'injections de produit de contraste.

Chapitre 4

Estimation du mouvement en situation de bi-transparence distribuée

Nous avons présenté dans le chapitre 3 une méthode robuste d'estimation de mouvements transparents, adaptée aux images cliniques bruitées et peu contrastées issues d'examen fluoroscopiques. Cependant, la méthode définie se limite au cas de la *bi-transparence*, c'est-à-dire aux images contenant exactement deux couches transparentes en mouvement. Le cas général est bien sûr plus complexe, mais l'observation de séquences cliniques révèle que des transparences faisant intervenir plus de deux couches en un point sont rares.

Cette remarque est développée dans la section 4.1, ce qui aboutit à la définition de la notion de *bi-transparence distribuée*. Ce terme se réfère aux images contenant un nombre arbitraire de couches, mais présentant en chaque pixel au plus deux couches transparentes. Nous insistons dans cette section sur le fait qu'une très large majorité des séquences cliniques se trouve dans cette configuration, et montrerons que les méthodes d'estimation que nous développons se concentrent naturellement sur les couches dominantes en cas de transparence d'ordre supérieur.

Ce chapitre et le suivant proposent des estimateurs adaptés à ce type de situation. Nous faisons pour le moment l'hypothèse *que la segmentation de l'image en ses différentes couches nous est donnée*. Le chapitre 5 proposera par contre une méthode ne nécessitant aucun *a priori* : elle comprendra l'estimation jointe du mouvement et de la segmentation en situation de bi-transparence distribuée.

Nous présentons à la section 4.2 un estimateur de mouvement dédié à une situation de bi-transparence distribuée à segmentation connue. Nous verrons que si l'estimation paramétrique des mouvements peut être directement dérivée de la méthode développée pour la configuration de bi-transparence dans le chapitre 3, l'initialisation de la méthode doit être revue.

Des résultats sur séquences synthétiques, et sur images cliniques réelles pré-segmentées, sont décrits à la section 4.3.

4.1 La bi-transparence distribuée

4.1.1 Définition

Nous fournissons une segmentation manuelle d'une image issue d'un examen cardiaque en ses différentes régions à deux couches à la Fig.4.1. Elle permettra d'illustrer notre propos. Les poumons, le coeur, le diaphragme et la colonne vertébrale y sont visibles. L'image peut être segmentée en deux zones mono-couche (en vert et en jaune), trois régions bi-couches et une petite zone présentant 3 couches (en gris). On peut se convaincre sans peine que cette dernière est peu significative, d'autant plus que les trois couches présentes y sont homogènes.

Nous désignons une telle configuration par le terme de **bi-transparence distribuée**.

Définition 2 *La bi-transparence distribuée se réfère à la classe des images contenant un nombre arbitraire de couches, mais ne présentant en chaque pixel qu'au plus deux couches transparentes.*

Nous avons tiré de nombreuses observations d'examens réels la conviction qu'une très large majorité de séquences cliniques est en situation de bi-transparence distribuée, puisque les transparences triples significatives sont rares. En toute rigueur bien sûr, la configuration est bien plus complexe puisque nous observons la projection d'organes 3D en mouvement. Néanmoins, un observateur peut segmenter dans la plupart des cas l'image de manière satisfaisante en zones à deux couches transparentes.

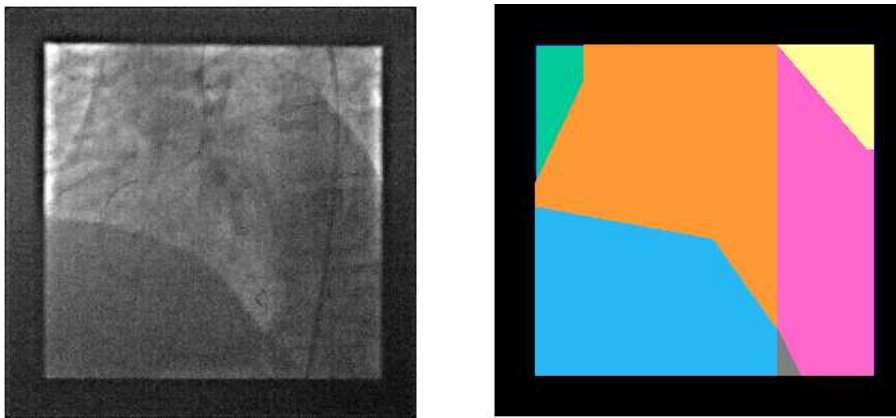


FIG. 4.1: À gauche une image de séquence de fluoroscopie, à droite sa segmentation manuelle en différentes zones à une ou deux couches transparentes (à l'exception de la petite région grise qui en compte 3). Une estimation de mouvement pour une configuration bi-transparente sur chacune de ces régions peut permettre de calculer les mouvements présents dans toute l'image.

4.1.2 Une segmentation en couches

Nous supposons dans ce chapitre que la segmentation de l'image en ses différentes couches est connue. Elle peut provenir soit de l'utilisateur (une interface permettant d'effectuer ce travail est présentée dans la partie 4.3), soit d'une segmentation automatique telle que celle proposée dans le chapitre 5.

Nous pouvons facilement en déduire la partition de l'image en ses régions bi-transparentes, et estimer les mouvements des couches sur ces régions en y appliquant la méthode présentée au chapitre 3. En suivant cette approche, nous pourrions bien estimer le mouvement de tous les organes, mais nous courrions le risque d'estimer plusieurs fois le mouvement de chaque organe, s'il est présent en différentes régions comme c'est le cas à la Fig.4.2. Par exemple, le coeur se trouve dans quatre régions "bitransparentes"

De fait, nous pouvons considérer la segmentation de l'image de deux façons (Fig.4.2) :

- en différentes *régions*, une pour chaque configuration de bi-transparence. Une région "bi-transparente" correspond alors à un couple de couches. Cependant, nous ne prenons alors pas en compte le fait qu'une couche puisse être présente sur plusieurs régions.
- en différentes *couches*, ce qui correspond mieux à la réalité physique du problème, mais oblige à manipuler une carte de segmentation nouvelle dont les unités (les couches) ne sont pas disjointes. (Il ne s'agit plus d'une partition de l'image, les couches peuvent se chevaucher puisqu'il y a transparence répartie sur l'image).

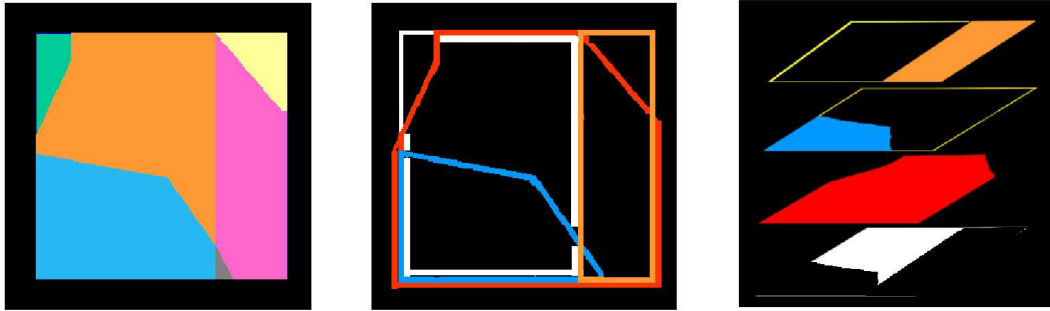


FIG. 4.2: À gauche, la segmentation de l'image clinique de la Fig.4.1 en ses différentes régions "bi-transparentes" (ou couples de couches). Au centre et à droite, la segmentation de l'image en ses différentes couches. Les couleurs sont utilisées indépendamment sur les cartes de segmentation en régions et en couches.

Sur la segmentation en couches de la Fig.4.2, quatre couches sont visibles : le coeur délimité en rouge, le diaphragme en bleu, la colonne vertébrale en orange et les poumons en blanc. Nous optons pour la *segmentation en couches* puisqu'elle modélise plus exactement le problème physique. Nous associons à chaque pixel \mathbf{p} une étiquette constituée par le couple $\{e_{\mathbf{p}}(1), e_{\mathbf{p}}(2)\}$ correspondant aux numéros des deux couches qui s'y trouvent. Sous l'hypothèse habituelle de mouvement constant sur deux intervalles de temps, la contrainte (2.39) est vérifiée en \mathbf{p} :

$$\begin{aligned}
 r(\mathbf{p}, \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}), \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p})) &= I(\mathbf{p}, t + 1) - I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}), t) - I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t) \\
 &\quad + I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t - 1) = 0 \quad (4.1)
 \end{aligned}$$

avec

$$\begin{cases} u_{\theta_{e_{\mathbf{p}}(1)}}(x, y) = a_{1,e_{\mathbf{p}}(1)} + a_{2,e_{\mathbf{p}}(1)}x + a_{3,e_{\mathbf{p}}(1)}y \\ v_{\theta_{e_{\mathbf{p}}(1)}}(x, y) = a_{4,e_{\mathbf{p}}(1)} + a_{5,e_{\mathbf{p}}(1)}x + a_{6,e_{\mathbf{p}}(1)}y \\ u_{\theta_{e_{\mathbf{p}}(2)}}(x, y) = a_{1,e_{\mathbf{p}}(2)} + a_{2,e_{\mathbf{p}}(2)}x + a_{3,e_{\mathbf{p}}(2)}y \\ v_{\theta_{e_{\mathbf{p}}(2)}}(x, y) = a_{4,e_{\mathbf{p}}(2)} + a_{5,e_{\mathbf{p}}(2)}x + a_{6,e_{\mathbf{p}}(2)}y. \end{cases} \quad (4.2)$$

Nous notons $\theta_k = (a_{1,k}, a_{2,k}, a_{3,k}, a_{4,k}, a_{5,k}, a_{6,k})$ les paramètres affines de mouvement correspondant à la couche k , et $\Theta = (\theta_1, \dots, \theta_K)$ la concaténation des paramètres de mouvement des K couches de l'image.

4.2 Estimation unifiée des mouvements des couches présentes dans l'image.

4.2.1 Résolution pratique du problème

Pour estimer les mouvements des couches, il faut déterminer Θ . Nous nous appuyons pour ce faire sur la contrainte (4.1). Nous cherchons donc à minimiser :

$$J(\Theta) = \sum_{\mathbf{p} \in \mathfrak{S}} r(\mathbf{p}, \theta_{e_{\mathbf{p}}(1)}, \theta_{e_{\mathbf{p}}(2)})^2 \quad (4.3)$$

où r est donné par (4.1), et où les vitesses sont liées aux paramètres par les équations (4.2). En fait, nous privilégions une approche robuste :

$$\min_{\Theta} \tilde{J}(\Theta) = \sum_{\mathbf{p} \in \mathfrak{S}} \rho(r(\mathbf{p}, \theta_{e_{\mathbf{p}}(1)}, \theta_{e_{\mathbf{p}}(2)})). \quad (4.4)$$

Il est encore plus important d'utiliser des estimateurs robustes dans le cas de la bi-transparence distribuée que dans celui de la bi-transparence simple puisque, outre les outils interventionnels, nous voulons ignorer les mouvements parasites issus d'une segmentation approximative de l'image en ses couches.

La fonction \tilde{J} est définie dans un espace à $6K$ paramètres. Elle est similaire à la fonction globale introduite dans le chapitre 3 et, de fait, nous la minimiserons de la même façon. Nous résumons rapidement la façon de résoudre ce problème.

Tout d'abord, il est possible de se ramener à la minimisation par les moindres carrés pondérés itérés :

$$\min_{\Theta} \tilde{J}^j(\Theta) = \sum_{\mathbf{p} \in \mathfrak{S}} w_j(\mathbf{p}) r(\mathbf{p}, \Theta)^2 \quad (4.5)$$

$$\text{avec } w_j(\mathbf{p}) = \frac{\rho'(r(\mathbf{p}, \Theta^{j-1}))}{2 \cdot r(\mathbf{p}, \Theta^{j-1})} \quad (4.6)$$

où Θ^{j-1} est l'estimé de Θ obtenu à l'itération précédente.

À chaque itération des moindres carrés pondérés, nous utilisons un formalisme incrémental de type Gauss-Newton dans un cadre multi-résolution. Nous initialisons les paramètres à une résolution donnée à partir des estimés obtenus au niveau précédent de la pyramide. Pour le niveau de résolution le plus grossier, nous adoptons l'initialisation développée plus

bas. Nous affinons alors itérativement l'estimation en minimisant une version linéarisée de (4.5).

Nous choisissons de nous appuyer sur une linéarisation de (4.5) *au second ordre* :

$$J_{\Theta}^j(\Theta + \delta\Theta) = \sum_{\mathbf{p} \in \mathfrak{S}} w_j(\mathbf{p}) (r(\mathbf{p}, \Theta + \delta\Theta) \nabla_{\Theta} r(\mathbf{p}, \Theta) + \frac{1}{2} \delta\Theta^T H_{\Theta} r(\mathbf{p}, \Theta) \delta\Theta)^2 \quad (4.7)$$

que nous minimisons itérativement en faisant appel à la méthode des gradients conjugués. Ces différents calculs sont développés en annexe B.

La segmentation par les couches mène à une approche unifiée permettant d'estimer les mouvements de chaque couche seulement une seule fois, et également d'estimer les mouvements de *toutes* les couches en même temps. Nous évitons ainsi les problèmes inhérents aux approches itératives : par quelle couche commencer ?, faut-il reboucler sur les différentes estimations ?, quand stopper ?, etc.

4.2.2 Initialisation

La méthode d'initialisation est différente de celle exposée au chapitre 3. Si nous gardons l'idée d'appliquer une mise en correspondance par blocs dans un premier temps, nous ne pouvons plus en extraire les modèles affines simplifiés grâce à une transformée de Hough globale pour les différentes couches présentes dans l'image.

En effet, une transformée de Hough appliquée à toute l'image n'utiliserait pas l'information de segmentation. On risquerait alors, en extrayant les pics les plus marqués dans la matrice d'accumulation, de sur-segmenter certaines couches tout en ignorant certaines autres. Une approche itérative, qui extrairait les modèles affines les uns après les autres en utilisant une transformée de Hough appliquée au seul support de la couche considérée, présente le même genre de risques.

Nous proposons une autre approche. Nous procédons en deux temps :

- nous estimons deux modèles affines de mouvement *sur chaque région "bitransparente"*.
- en nous basant sur la segmentation de l'image en couches, nous attribuons chacun des modèles de mouvement estimés au niveau des régions bitransparentes à la couche qui lui correspond.

Au terme de cette étape, chaque couche se voit attribuer un ou plusieurs modèle de mouvement desquelles nous extrayons un modèle affine de mouvement unique.

La méthode proposée est illustrée sur la figure 4.3. La nouveauté porte sur la transition "modèles de mouvement estimés sur les régions bitransparentes" à "modèles estimés sur les couches", qui est moins directe qu'il n'y paraît pour deux raisons :

- Nous calculons sur chaque région transparente R_{kl} (intersection des couches C_k et C_l) *deux* modèles affines $\{\theta_{\{k,l\}}^1, \theta_{\{k,l\}}^2\}$. Mais nous ne savons pas lequel correspond à la couche k et lequel correspond à la couche l .
- Chaque couche C_k peut s'étendre sur plusieurs (m) régions. Il nous faudra cependant extraire un modèle unique des m modèles de mouvement attribués à C_k sur ces m régions.

Sur l'exemple de la Fig.4.3, les champs de déplacement représentés en bleu, rouge et cyan, estimés sur les régions bi-transparentes (figure du centre), correspondent en fait au mouvement de la couche rouge (figure de gauche). Il faut savoir les lui attribuer, même quand les champs de déplacement estimés sont plus bruités que sur cette figure.

En résumé, l'estimation du mouvement par régions calcule deux modèles affines

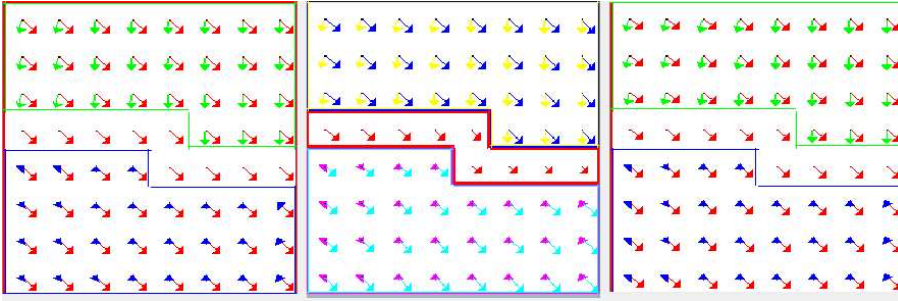


FIG. 4.3: Illustration de la méthode d'initialisation de l'estimation du mouvement transparent en situation de bi-transparence distribuée à segmentation donnée. À gauche la vérité-terrain, au centre les deux modèles de mouvement calculés sur chaque région bi-transparente (avec la méthode du chapitre 3), à droite les modèles de mouvement qu'ils permettent d'estimer pour chaque couche. Le choix des couleurs est indifférent d'une figure à l'autre.

$\{\theta_{\{k,l\}}^1, \theta_{\{k,l\}}^2\}$, qu'il nous faut attribuer aux couches k et l . Nous noterons les modèles ainsi attribués $\theta_{\{k,l\},k}$ et $\theta_{\{k,l\},l}$ respectivement. D'autre part, il nous faudra ensuite, sur chaque couche, calculer le modèle de mouvement unique $\hat{\theta}_k$ qui corresponde le mieux aux $\theta_{\{k,l\},k}$. Ce raisonnement est schématisé à la figure 4.4.

En fait, les deux dernières étapes sont liées puisque l'attribution des $\{\theta_{\{k,l\}}^1, \theta_{\{k,l\}}^2\}$ à chaque

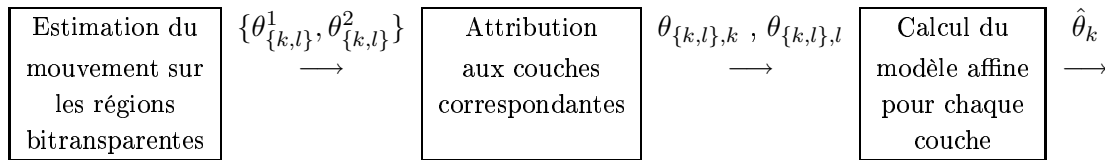


FIG. 4.4: Approche suivie lors de l'initialisation

couche sera améliorée par une bonne estimation des mouvements de chaque couche, et réciproquement. Nous optons donc pour le schéma itératif présenté à la figure 4.5, qui fait également intervenir une initialisation de l'appariement.

Les quatre modules de traitement intervenant dans ce schéma font l'objet des sous-sections suivantes.

4.2.2.1 Estimation des modèles de mouvement sur chaque région

Nous proposons deux approches pour estimer les modèles affines de mouvement présents sur chaque région bi-transparente de l'image, selon que l'on privilégie la précision de calcul ou le temps de traitement :

- nous pouvons effectuer une estimation du mouvement en situation de bi-transparence simple sur chaque région en suivant l'intégralité de la méthode du chapitre 3,
- ou bien nous pouvons utiliser seulement l'initialisation développée en 3.4.3.2, qui consiste alors à utiliser la transformée de Hough *au niveau de chaque région bi-*

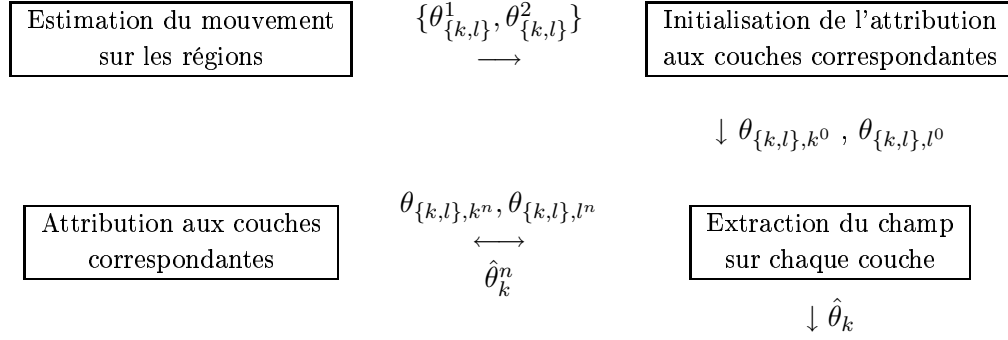


FIG. 4.5: Approche pratique suivie lors de l'initialisation

transparente. Cette alternative est plus rapide mais moins précise que l'utilisation intégrale de la méthode du chapitre 3. Cela peut néanmoins former un compromis intéressant puisque nous ne recherchons à ce stade qu'une initialisation de la méthode d'estimation en situation de bi-transparence distribuée.

Ces estimations associent à chaque couple de couches k et l (définissant une région bi-transparente) deux modèles affines $\theta_{\{k,l\}}^1$ et $\theta_{\{k,l\}}^2$. Par contre, nous ne savons pas lequel de ces deux modèles correspond au mouvement de la couche k (respectivement l).

4.2.2.2 Attribution des modèles affines estimés sur les régions aux couches

Lorsqu'une estimation des modèles de mouvement $\hat{\theta}_k$ de chacune des couches est déjà connue, on peut les associer aux modèles affines estimés sur chaque région, de sorte à ce que la distance

$$\sum_{\mathbf{p} \in \mathfrak{S}} \|\mathbf{w}_{\theta_{\{e_1(\mathbf{p}), e_2(\mathbf{p})\}, e_1(\mathbf{p})}}(\mathbf{p}) - \mathbf{w}_{\theta_{e_1(\mathbf{p})}}(\mathbf{p})\|^2 + \|\mathbf{w}_{\theta_{\{e_1(\mathbf{p}), e_2(\mathbf{p})\}, e_2(\mathbf{p})}}(\mathbf{p}) - \mathbf{w}_{\theta_{e_2(\mathbf{p})}}(\mathbf{p})\|^2 \quad (4.8)$$

soit minimale.

En d'autres termes, on choisit sur chaque région celui des deux appariements pour lequel

- les champs de déplacement issus des modèles affines estimés sur la région
- et les champs de déplacement issus des modèles affines calculés sur chaque couche

sont les plus proches.

4.2.2.3 Extraction des modèles à étiquetage connu

Lorsque les appariements sont connus sur chaque région, on sait attribuer à chaque couche k les modèles affines qui lui correspondent sur chaque région de son support. Il faut maintenant estimer un modèle affine $\hat{\theta}_k$ "unifié" qui corresponde le mieux possible aux $\theta_{\{k,l\},k}$.

Nous souhaitons là aussi minimiser sur la couche la distance entre les champs de déplacement issus de ces différents modèles sur la couche C_k . On veut minimiser par rapport à $\hat{\theta}_k$

le critère :

$$F_k(\theta_k) = \sum_{\mathbf{p} \in C_k} \|\mathbf{w}_{\theta_{\{k,l\},k}} - \mathbf{w}_{\theta_k}\|^2 \quad (4.9)$$

$$= \sum_{\mathbf{p} \in C_k} (\mathbf{B}(\mathbf{p})\theta_{\{k,l\},k} - \mathbf{B}(\mathbf{p})\theta_k)^T (\mathbf{B}(\mathbf{p})\theta_{\{k,l\},k} - \mathbf{B}(\mathbf{p})\theta_k) \quad (4.10)$$

où

$$\mathbf{B}(\mathbf{p}) = \begin{bmatrix} 1 & x & y & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x & y \end{bmatrix} \quad (4.11)$$

en exploitant le fait que les mouvements recherchés suivent un modèle affine complet. L'équation (4.10) est un critère quadratique classique, dont le minimum est :

$$\hat{\theta}_k = \left(\sum_{\mathbf{p} \in C_k} \mathbf{B}(\mathbf{p})\mathbf{B}^T(\mathbf{p}) \right)^{-1} \sum_l \left(\sum_{\mathbf{p} \in R_{k,l}} \mathbf{B}(\mathbf{p})\mathbf{B}(\mathbf{p})^T \theta_{\{k,l\},k} \right). \quad (4.12)$$

Nous avons noté $R_{k,l}$ la région de l'image où les couches C_k et C_l sont présentes.

En pratique, pour pallier d'éventuelles mauvaises estimations sur certaines régions, nous cherchons à minimiser la mesure robuste suivante :

$$F_k(\theta_k) = \sum_{\mathbf{p} \in C_k} \rho(\|\mathbf{w}_{\theta_{\{k,l\},k}} - \mathbf{w}_{\theta_k}\|, C) \quad (4.13)$$

où $\rho(\cdot, C)$ désigne la fonction de Tukey. C est le facteur d'échelle, que l'on peut considérer comme le seuil de tolérance sur les distances entre les déplacements permettant d'identifier un outlier ; nous le fixons à 2.

Nous utilisons la méthode des moindres carrés pondérés, en minimisant itérativement :

$$F_k(\theta_k) = \sum_{\mathbf{p} \in C_k} w(\mathbf{p}) \|\mathbf{w}_{\theta_{\{k,l\},k}} - \mathbf{w}_{\theta_k}\|^2 \quad (4.14)$$

$$= \sum_{\mathbf{p} \in C_k} w(\mathbf{p}) (\mathbf{B}(\mathbf{p})\theta_{\{k,l\},k} - \mathbf{B}(\mathbf{p})\theta_k)^T (\mathbf{B}(\mathbf{p})\theta_{\{k,l\},k} - \mathbf{B}(\mathbf{p})\theta_k) \quad (4.15)$$

et en remettant à jour les poids :

$$w(\mathbf{p}) = \frac{\rho'(\|\mathbf{w}_{\theta_{\{k,l\},k}} - \mathbf{w}_{\hat{\theta}_k}\|^2, C)}{2\|\mathbf{w}_{\theta_{\{k,l\},k}} - \mathbf{w}_{\hat{\theta}_k}\|^2}. \quad (4.16)$$

À chaque itération, le minimum est donné par :

$$\hat{\theta}_k = \left(\sum_{\mathbf{p} \in C_k} w(\mathbf{p})\mathbf{B}(\mathbf{p})\mathbf{B}^T(\mathbf{p}) \right)^{-1} \sum_l \left(\sum_{\mathbf{p} \in R_{k,l}} w(\mathbf{p})\mathbf{B}(\mathbf{p})\mathbf{B}(\mathbf{p})^T \theta_{\{k,l\},k} \right). \quad (4.17)$$

4.2.2.4 Initialisation de l'appariement

Il nous faut encore initialiser au mieux cet algorithme itératif. Nous proposons pour ce faire de considérer les couches les unes après les autres, en commençant par la plus grande. Chaque couche ne donnant en pratique naissance qu'à un petit nombre de régions bi-transparentes (de 1 à 4 en général), nous choisissons de rechercher les modèles à associer à la couche considérée en testant de manière exhaustive toutes les combinaisons possibles. Nous procédons itérativement. Pour chaque couche C_k , et chaque scénario d'attribution f des modèles de mouvement estimés sur les régions $R_{k,l}$ à C_k

$$f : \{\theta_{\{k,l\}}^1, \theta_{\{k,l\}}^2\}_{l \in [1,n]} \longrightarrow \theta_{\{k,l\},k} \quad (4.18)$$

- nous calculons le modèle affine $\hat{\theta}_k$ de C_k en utilisant (4.12). Nous préférons ne pas utiliser l'équivalent robuste de cette formule à ce stade.
- puis, nous calculons la distance sur le support de C_k
 - entre les modèles estimés $\theta_{\{k,l\},k}$ sur les régions $R_{k,l}$ et attribués à C_k
 - et le modèle estimé $\hat{\theta}_k$ sur C_k :

$$\sum_{\mathbf{p} \in C_k} \|\mathbf{w}_{\theta_{\{k,l\},k}}(\mathbf{p}) - \mathbf{w}_{\hat{\theta}_k}(\mathbf{p})\|^2. \quad (4.19)$$

Nous gardons au final le scénario d'attribution des modèles de mouvement (estimés au niveau des régions) donnant le critère (4.19) le plus faible. Les modèles affines "régionaux" ainsi attribués n'entrent plus en ligne de compte pour le traitement des couches suivantes. Une telle stratégie exhaustive est envisageable d'une part parce que le nombre de scénarios est peu élevé, et d'autre part parce que les tests à effectuer pour chaque scénario sont très rapides. Nous travaillons en effet à ce stade au niveau des modèles de mouvement et non à celui des images. Il n'est donc pas question d'interpolations. Cela dit, ces tests systématiques ne fournissent qu'une première hypothèse d'appariement puisque les couches sont traitées séquentiellement et non globalement.

L'algorithme 2 résume la méthode d'initialisation.

4.3 Résultats

Nous présentons dans cette partie des résultats sur des séquences d'images synthétiques, qui nous permettent de quantifier le fonctionnement de l'estimateur proposé, ainsi que sur des exemples de séquences cliniques réelles, segmentées manuellement.

4.3.1 Résultats sur séquences synthétiques

4.3.1.1 Synthèse d'images réalistes en situation de bi-transparence distribuée

Il nous faut aménager notre façon de générer des images réalistes pour pouvoir en produire en situation de bi-transparence *distribuée*. La méthode décrite en 1.3.2 n'est en effet adaptée qu'à la bitransparence simple : les couches sont présentes sur l'intégralité de l'image. Désormais, nous voulons combiner un nombre donné de couches, à mouvements *et segmentation* connus. En pratique, nous allons considérer le cas d'une couche présente

Algorithme 2 Algorithme d'initialisation de l'estimation de mouvement transparent en situation de bi-transparence distribuée à segmentation donnée .

- Estimation du mouvement sur les régions.
 - Mise en correspondance par blocs,
 - Transformée de Hough,
 - Éventuellement estimation paramétrique.

- Initialisation de l'attribution aux couches correspondantes.

Tests exhaustifs :

- Calcul des modèles de mouvement par couches en utilisant (4.12),
 - Évaluation de l'erreur d'appariement.
-
- Itérativement
 1. Attribution aux couches correspondantes
(Choix de la combinaison menant à la plus petite erreur d'appariement).
 2. Extraction du champ sur chaque couche (par formule (4.17)).
-

partout qui s'ajoute à deux autres couches partitionnant l'image (nous représentons un exemple de champs de mouvement dans cette configuration à la figure 4.6).

Cette notion de segmentation est à prendre avec précaution car il n'y a pas dans la réalité clinique de sauts abrupts dans les mouvements de deux couches adjacentes, menant à des segmentations nettes. Ainsi les poumons sont-ils entraînés par les mouvements du cœur dans son voisinage immédiat, tout comme ils sont entraînés par le diaphragme dans les régions qui lui sont proches.

Puisqu'il est délicat de définir une limite claire entre deux couches adjacentes, nous séparons dans nos simulations les deux couches qui partitionnent l'image par une *zone de transition* qui permet de passer de manière souple d'un mouvement à l'autre (Fig.4.6).

Nous n'introduisons pas d'informations relatives à l'existence de frontières de couches dans les niveaux de gris des images car nous avons pris le parti de définir les couches sur la base de leur mouvement seulement.

En pratique, nous considérons deux images cliniques I_1 et I_2 , et segmentons la deuxième en deux couches (avec une zone de transition). Nous associons à chacune des trois couches ainsi définies (une pour I_1 , et deux pour I_2) un modèle de mouvement $\theta_i, i = 1...3$.

Nous calculons alors pour chaque image les déplacements correspondants en chaque pixel, sur la base des modèles de mouvement affines donnés et de la partition en couches. Les déplacements dans la zone de transition sont obtenus en pondérant les modèles affines des deux couches voisines par la distance à chacune d'entre elles.

Nous sommes alors en mesure d'interpoler la valeur de chacune des couches I_i au cours de la séquence, que nous combinons ensuite comme en 1.3.2 pour obtenir des images réalistes

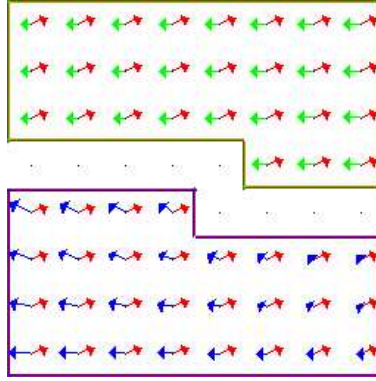


FIG. 4.6: Configuration typique des images synthétiques générées : une couche abdominale est en translation sur toute l'image (champ rouge), et une couche cardiaque subit deux champs de mouvements différents (champs bleu et vert). Ils sont séparés par une zone de transition.

en situation de bi-transparence distribuée, et à vérité-terrain connue : mouvements, segmentation, bruit, diffusé...

Dans le cadre de nos simulations, nous utilisons une image issue d'un examen abdominal I_1 , à laquelle nous associons un mouvement de translation, et une seconde image cardiaque I_2 segmentée en deux couches de surface égale (et séparées par une zone de transition d'épaisseur un bloc 32×32), que nous animons de mouvements affines simplifiés perturbés différents (Fig.4.6). (On autorise une variation des coefficients affines purs de 20% autour du modèle simplifié, comme en 3.5).

Comme dans le chapitre 3, chacune des simulations fait l'objet de 250 générations aléatoires de mouvements (ayant une amplitude d'au plus 8 pixels, et distants d'au moins 2 pixels sur leur support respectif). Lorsqu'il n'est pas précisé, le diffusé est de 20%. Notre métrique pour juger de la performance de l'estimation de mouvement est la même que dans le chapitre 3, à une nuance près : il s'agit de la distance (en pixels) entre le déplacement estimé et le déplacement réel, moyennée sur toute l'image *hors les zones de transition*.

Nous savons en effet que les tissus y suivent des mouvements transitoires mal déterminés, qu'il est illusoire dans la pratique d'espérer les calculer avec précision puisque l'observateur lui-même ne sait pas comment segmenter précisément la zone en question. Seules des méthodes d'estimation de champs denses pourraient éventuellement prétendre à calculer ces déplacements, mais comme discuté précédemment, ces approches ont un coût de calcul élevé et surtout présentent généralement une grande sensibilité au bruit.

4.3.1.2 *Expériences sur séquences synthétiques*

Nous avons repris les prétraitements retenus au chapitre 3 : la MTF est inversée, un filtre gaussien d'écart-type 1,5 est appliqué et enfin les images sont soumises à une fonction d'encodage en racine carrée (commençant par une partie linéaire).

Nous rassemblons sur le tableau 4.1 les erreurs d'estimation moyennes effectuées (en pixel) pour des générations de séquences d'images de différents types : graphie (examen diagnos-

tique, $\sigma = 10$) et fluoroscopie (examen interventionnel, $\sigma = 20$), taux de diffusé de 0%, 20% et 50%. Nous utilisons deux versions de l'estimateur :

- dans la première, les modèles de mouvement sont calculés sur chaque région par une mise en correspondance par blocs suivie de la transformée de Hough, mais sans estimation paramétrique ;
- dans la seconde, la méthode d'estimation de mouvements transparents reprend l'intégralité de l'algorithme du chapitre 3.

TAB. 4.1: Tests sur séquences synthétiques en situation de bi-transparence distribuée à segmentation connue. Les images de graphie sont entachées d'un bruit d'écart-type $\sigma = 10$, les images fluoroscopiques d'un bruit d'écart-type $\sigma = 20$.

Type d'examen	Calcul par Hough		Estimation paramétrique	
	Erreur d'estimation	Temps de calcul	Erreur d'estimation	Temps de calcul
Graphie sans diffusé	0,42	29	0,42	60
Graphie diffusé 20%	0,70	40	0,70	69
Graphie diffusé 50%	0,92	41	0,95	75
Fluoroscopie sans diffusé	1,49	45	1,68	84
Fluoroscopie diffusé 20%	1,81	79	1,83	123
Fluoroscopie diffusé 50%	3,46	46	3,54	88

Les estimés de mouvement sur images diagnostiques sont très satisfaisants. Leur précision est sub-pixelique, de l'ordre de 0.7 pixels dans le cas le plus représentatif d'un diffusé de 20%. La précision reste intéressante sur images fluoroscopiques, à part dans le cas où le diffusé est très élevé. Une valeur typique de la précision moyenne est 1.8 pixels.

Les résultats sont cependant moins bons que dans le cas de la bi-transparence simple. Nous y voyons deux raisons majeures :

- Les estimations bi-transparentes initiales par régions portent sur des zones deux fois plus petites que dans le chapitre 3, et y sont donc moins fiables, ce qui explique un certain nombre d'erreurs.
- La contrainte que nous imposons dans nos tests, à savoir "les mouvements générés doivent être distants de plus de 2 pixels en moyenne sur chacune des couches concernées" est moins forte que dans le cas de deux couches sur toute l'image. Les mouvements étant moins fortement divergents sur une région particulière que sur l'image dans son ensemble, deux mouvements "proches" au sens de ce critère (distants de 2 pixels en moyenne) ne sont pas seulement proches *en moyenne* mais deviennent proches *partout sur la région considérée*.

À l'inverse, dans le cas de la bi-transparence simple, des mouvements "proches" (au sens du critère présenté ci-dessus) étaient dans la pratique souvent assez nettement différents sur une région délimitée de l'image. Cela permettait la plupart du temps d'amorcer l'estimation du modèle de mouvement correspondant.

Nous présentons un exemple d'une situation de génération de mouvements proches à la Fig.4.7. Dans ce cas particulier, les *trois* mouvements sont proches. Les estimations bi-transparentes sur les deux régions permettent d'estimer deux des

trois modèles de mouvement correctement, mais fournissent un troisième modèle totalement erroné, de sorte que l'estimation paramétrique globale au niveau de l'image ne permet pas de le retrouver.

Nous insistons sur le fait que des situations telles que celles de la figure 4.7 apparaissent beaucoup plus fréquemment dans les configurations bi-transparentes distribuées présentées que sur des séquences en bi-transparence simple. Les champs bleus et verts de la vérité-terrain seraient plus nettement divergents si on les prolongeait sur l'image entière.

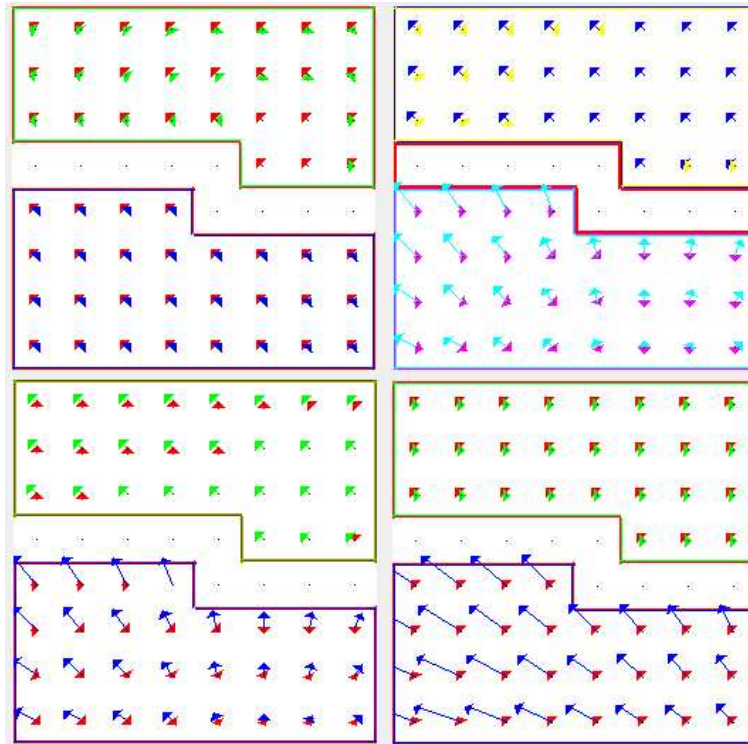


FIG. 4.7: Exemple d'une mauvaise estimation. En haut à gauche la vérité-terrain. En haut à droite les champs issus des modèles de mouvement estimés sur chaque région. En bas à gauche les mêmes champs attribués aux couches. En bas à droite les modèles de mouvement finaux obtenus au terme de l'estimation paramétrique.

Ce phénomène est d'autant plus fréquent que les images sont difficiles (du fait du bruit et du diffusé) et que les déplacements réels sont proches. Le diagramme à droite de la Fig.4.8 trace pour chaque simulation les erreurs d'estimation commises, en fonction de la plus petite distance entre deux des trois champs de mouvement issus des modèles de la vérité terrain. Il montre que l'estimation est plus difficile quand les champs sont proches. Nous présentons d'autre part sur la gauche de la Fig.4.8 les erreurs commises sur les images diagnostiques et interventionnelles avec un taux de diffusé de 20% pour chacun des 250 tests. Nous y remarquons une consistance plus faible que dans le cas de la bi-transparence simple, mais cela est dû aussi en grande partie au phénomène des champs fréquemment proches.

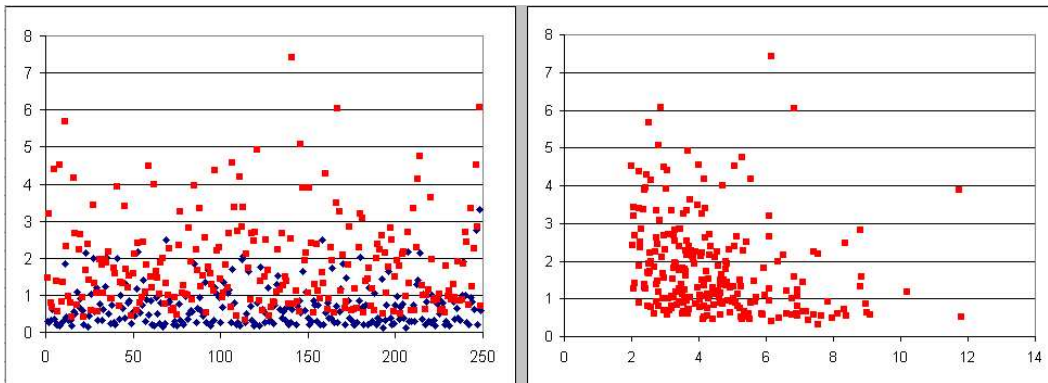


FIG. 4.8: À gauche, tracé des erreurs d'estimation commises pour chacun des 250 tests sur des images diagnostiques avec 20% de diffusé (en bleu), et des images interventionnelles avec 20% de diffusé (en rouge). À droite, le diagramme des erreurs d'estimation en fonction de la distance entre les deux champs les plus proches (sur des images interventionnelles avec 20% de diffusé).

D'autre part, le tableau 4.1 montre que les deux versions de la méthode d'estimation des mouvements transparents en situation de bi-transparence distribuée se différencient peu en terme de précision, de sorte que nous préférons effectuer l'estimation de mouvement par région bitransparente sans estimation paramétrique pour gagner en rapidité d'exécution.

Enfin, notons que le temps de calcul est fortement dépendant de la difficulté de la situation traitée. Si on arrive à convergence en 40 secondes environ sur des images de graphie, il faut compter deux fois plus de temps sur des images fluoroscopiques.

Nous étudions à la Tab.4.2 le comportement de la méthode d'estimation pour la bi-transparence distribuée lorsque l'hypothèse de constance temporelle des mouvements est relâchée. Pour ce faire, nous générons pour chaque couche deux modèles de mouvement proches simulant une variation temporelle de mouvement (comme expliqué au paragraphe 3.5.1), que nous appliquons sur des séquences d'images diagnostiques et interventionnelles (à 20% de diffusé).

Comme dans le chapitre 3, nous observons une dégradation progressive des performances de l'estimateur, ce qui montre que cette hypothèse admet des violations légères sans conséquences dramatiques pour la qualité des estimés.

4.3.2 Traitement de séquences réelles

4.3.2.1 Dispositif pratique

Afin que l'utilisateur puisse segmenter l'image en couches, nous avons développé l'interface présentée à la figure 4.9. Elle permet de déterminer successivement les polygones dans lesquels s'inscrivent les couches.

TAB. 4.2: Moyenne de l'erreur d'estimation pour différents niveaux de bruit et différentes variations temporelles des mouvements successifs (avec 20% de diffusé).

Bruit	Variation temporelle des mouvements successifs			
	0%	10%	20%	30%
Graphie $\sigma = 10$	0.70	0.81	0.90	0.99
Fluoroscopie $\sigma = 20$	1.81	2.12	2.28	2.49

Nous répercutons ensuite cette segmentation au niveau des blocs, en attribuant chacun d'entre eux aux deux couches qui y sont les plus présentes. Lorsque trois couches au moins occupent plus d'un tiers du bloc, nous l'ignorons lors de l'estimation de mouvement car nous considérons qu'il est en situation de tri-transparence.

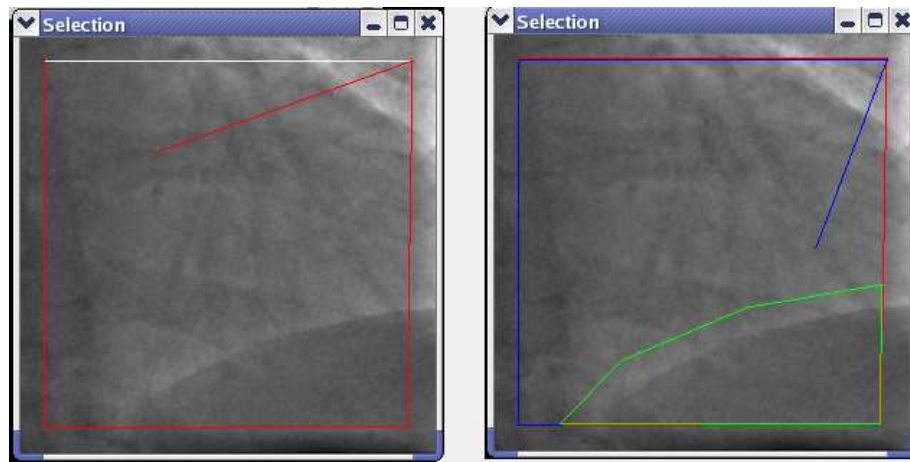


FIG. 4.9: Interface de détermination des couches.

4.3.2.2 Résultats

Généralités. Nous présentons les résultats obtenus sur trois séquences fluoroscopiques et une séquence de graphie avec injection de produit de contraste. Nous traçons dans ce document les modèles de mouvement estimés à quatre instants espacés jugés intéressants, avec l'image de la séquence correspondante, et nous donnons une évaluation visuelle de la qualité des estimations sur la séquence entière.

Le lecteur pourra mieux juger des performances de la méthode proposée sur une séquence entière en se reportant au site <http://www.irisa.fr/Vista/Equipe/People/Auvray/-DemoBiDistributedTransparencyWithKnownSegmentation.html>, qui présente pour chacune des quatre estimations la séquence de trois images qui lui correspond, ainsi qu'une vidéo de l'intégralité des champs issus des modèles de mouvement estimés le long de la séquence. Pour des raisons de lisibilité, nous avons multiplié la longueur des vecteurs représentés par 4. Rappelons également que les déplacements représentés sont, en conformité avec les notations de la thèse, *orientés vers le passé*.

Enfin, soulignons que la méthode d'estimation de mouvement en situation de bi-

transparence distribuée estime *deux* mouvements sur chaque région, même quand seul un mouvement est présent à cet instant sur la région en question. Cela se traduit par l'apparition d'un second modèle "fictif", sans réalité physique, qui gêne la lisibilité des tracés de champs de mouvement, présentés à certains instants. Nous verrons par contre que cela perturbe peu l'application visée de débruitage compensé en mouvement avec transparence.

Séquence 10. La première séquence traitée (à la figure 4.10) correspond à l'exemple de bi-transparence distribuée donné en début de chapitre à la figure 4.1. Il s'agit d'un examen fluoroscopique. Les images ont pour dimension 864×864 pixels. On y observe le cœur (masse gris clair, en forme de poire, centrée sur l'image), le diaphragme (masse sombre en bas à gauche), les tissus clairs des poumons, la colonne vertébrale (à droite) et les côtes. En outre, un cathéter et deux guides sont présents.

Nous avons segmenté l'image interactivement (Fig.4.2), en attribuant une couche au cœur, une au diaphragme, une à la colonne vertébrale et une aux poumons.

Sur les deux premières images présentées (Fig.4.10), les quatre mouvements des couches sont bien estimés : celui du cœur est représenté en vert, du diaphragme en bleu, de la colonne en jaune et des poumons en rouge.

Pour les deux images suivantes (Fig.4.10), les mouvements estimés sont également cohérents avec l'observation, mais attribués différemment. Ils se trouve qu'aux deux instants présentés, le cœur a un mouvement de faible amplitude, surtout à droite de l'image. Il correspond au modèle estimé dont le champ correspondant est tracé en rouge sur les poumons, en bleu sur le diaphragme... et en jaune sur la colonne puisqu'il se confond avec le mouvement du fond (statique). Le mouvement dominant devient donc à ces instants celui de la respiration, qui anime diaphragme comme poumons, et est représenté en vert... alors même que l'utilisateur destinait cette couche aux mouvements cardiaques.

Cet exemple illustre donc la faculté de la méthode à estimer correctement les mouvements de la séquence même quand la segmentation donnée n'est pas en conformité absolue avec les mouvements en présence. Ici, la confusion vient du fait que la région à droite de l'image met en fait en jeu trois couches : la colonne, le cœur et les poumons. Lorsque les mouvements des trois couches sont différents, celui des deux couches dominantes (le cœur et la colonne vertébrale) sont estimés. On peut par contre estimer le mouvement des poumons aux instants où le cœur est immobile.

Globalement, les mouvements estimés semblent corrects dans 95% des cas pour les couches autres que le diaphragme. Il est par contre difficile de quantifier la performance au niveau du diaphragme : il est si peu contrasté et si souvent statique que c'est fréquemment le mouvement du guide qui est estimé (et de manière assez peu fiable étant donné la rapidité de ses mouvements).

Séquence 20. La séquence présentée à la figure 4.11 est particulièrement intéressante dans la mesure où la structuration fournie par la méthode d'estimation du mouvement transparent apparaît plus pertinente que celle qu'un utilisateur pourrait faire sur la base de connaissances anatomiques. Cette séquence correspond à un examen cardiaque (fluoroscopique) sous une angulation complexe. Les images traitées ont pour dimensions 576×576 pixels.

La masse gris moyen correspond au cœur, celle qui est plus foncée au diaphragme, et les tissus clairs sont ceux des poumons. Les côtes sont également visibles. Nous observons en

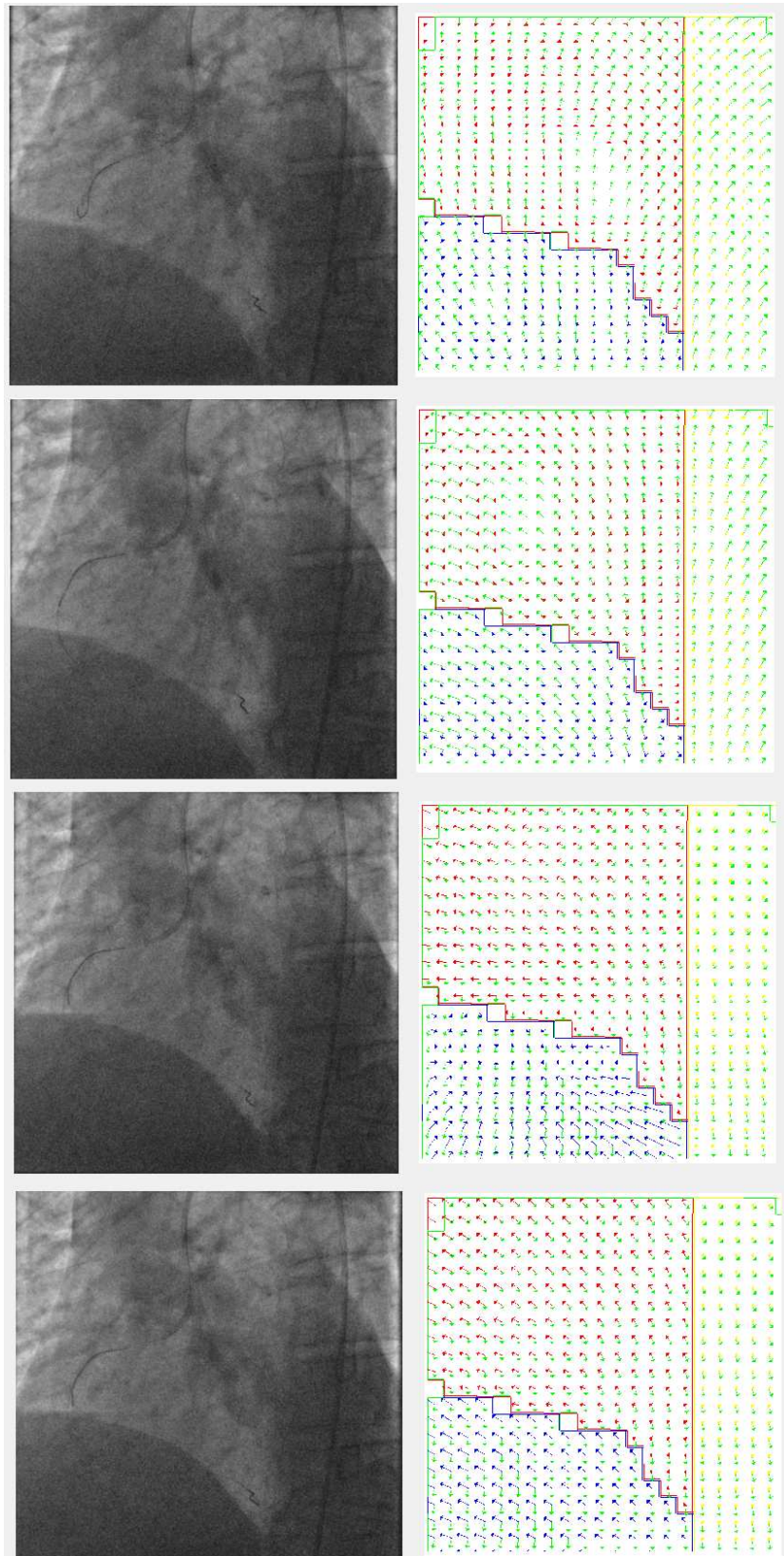


FIG. 4.10: Estimation du mouvement de la séquence fluoroscopique 10, en situation de bi-transparence distribuée (voir texte principal).

outre un guide dont l'extrémité radio-opaque est très visible, et des agrafes solidaires du cœur.

Nous avons segmenté interactivement l'image en dessinant quatre couches :

- la première s'étend sur toute l'image ;
- la seconde correspond au diaphragme ;
- la troisième au cœur ;
- et la dernière aux poumons.

Nous nous sommes appuyés pour ce faire sur des considérations anatomiques, et non pas sur des critères de mouvement, comme il est apparu par la suite. Notre interprétation était que la couche correspondant à l'ensemble de l'image devait être allouée à la couche statique, à travers certaines agrafes et les côtes qui semblaient immobiles, et que les trois autres devaient suivre les mouvements indépendants des organes auxquels ils correspondent.

En fait, une observation de la séquence au sens du mouvement révèle que nous sommes en situation de *transparence triple sur chacune des trois régions* bi-couches issue de la segmentation effectuée. Sur chacune d'entre elles sont présentes :

- une couche statique faiblement contrastée,
- une couche en mouvement correspondant aux battements cardiaques¹ (entraînant une partie des tissus des poumons),
- une couche correspondant au mouvement de respiration (associée au diaphragme et à une partie des poumons).

La couche statique étant moins contrastée que les deux autres, son mouvement n'est estimé que lorsqu'il rejoint celui de la couche "cardiaque" (sur la dernière image présentée ici). Les mouvements des deux couches dominantes sont estimés correctement sur les trois régions de l'image, bien que la segmentation fournie ne corresponde pas à une analyse exacte de l'image en terme de mouvements.

Sur l'intégralité de la séquence, en faisant abstraction des quelques mouvements "inventés" lorsque seul un mouvement apparent est présent dans la région bi-couche considérée, les mouvements estimés semblent qualitativement corrects dans 90% des cas environ.

Séquence 26. La séquence de la figure 4.12 présente la même configuration mais est plus difficile pour trois raisons. Des stents insérés dans les artères coronaires rendent leurs mouvements plus perceptibles et donc introduisent un mouvement parasite, un cathéter déplacé par le médecin fait son apparition dans le dernier tiers de la séquence, et un mouvement respiratoire de grande amplitude apparaît dans la seconde partie de la séquence. Pour ces raisons, on peut évaluer le taux de "bonne estimation" de 60% à 70% sur l'intégralité de la séquence (en ignorant les mouvements "fictifs" sur les zones mono-couches).

La première et la troisième image présentées reflètent correctement les trois mouvements en présence : battement cardiaque, mouvement du diaphragme et fond statique. La deuxième image se réfère à un instant où la frontière du cœur est immobile, mais où l'algorithme capte les déplacements de l'artère coronaire (vérifiant approximativement l'hypothèse de constance temporelle ici, ce qui n'est pas toujours le cas). Enfin, la dernière image correspond à un mouvement respiratoire de grande amplitude se formant dans la deuxième partie de la séquence, et qui est bien estimé ici.

¹De faible amplitude et à fréquence faible ici.

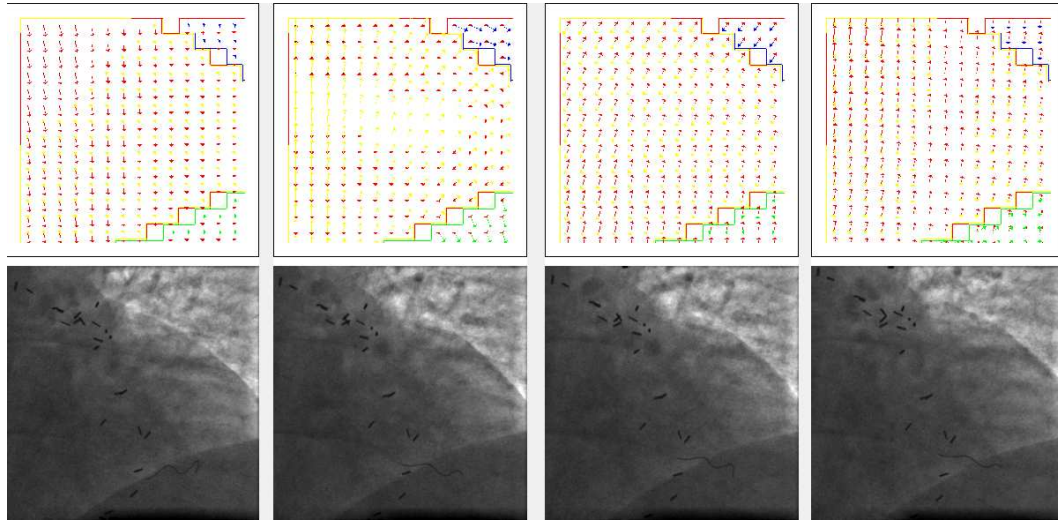


FIG. 4.11: Estimation du mouvement de la séquence fluoroscopique 20, en situation de bi-transparence distribuée (voir texte principal).

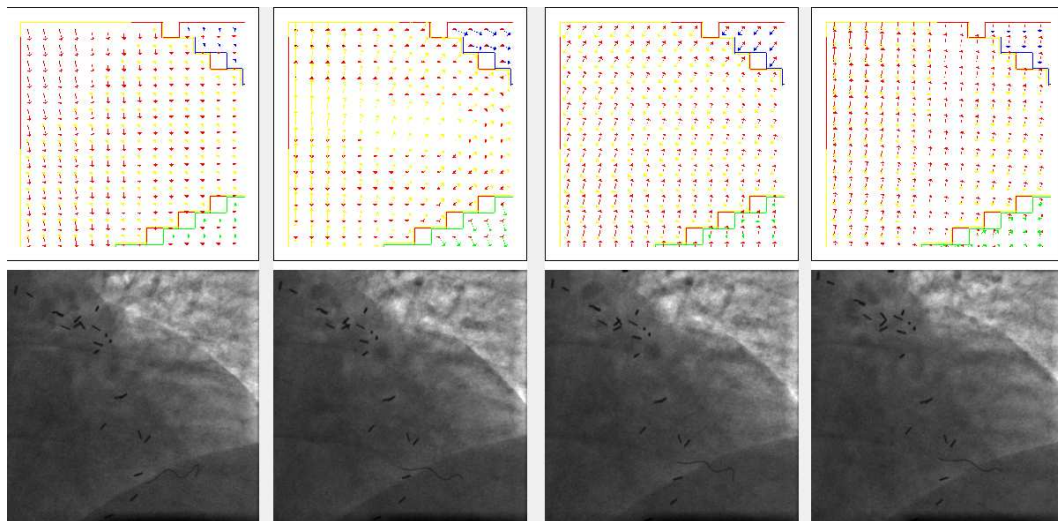


FIG. 4.12: Estimation du mouvement de la séquence fluoroscopique 26, en situation de bi-transparence distribuée (voir texte principal).

Séquence 33. Nous rapportons enfin les résultats obtenus sur une séquence diagnostique de taille 576×576 pixels (Fig.4.13). On y observe le cœur (masse gris clair, dans laquelle se trouvent un cathéter et un guide), les tissus clairs des poumons (entraînés par les mouvements du cœur) et le diaphragme (masse gris foncé) qui s'ajoutent à un fond

statique matérialisé par les côtes notamment. Nous avons opté pour une segmentation en trois couches : une pour le fond, une pour le cœur et la dernière pour le diaphragme.

Concernant la première image, une estimation pertinente du mouvement du fond (quasi-statique) et du cœur est fournie. Le mouvement ascendant du diaphragme ne correspond pas à la réalité clinique du fait du problème de l'observation généralisée.

Sur la seconde image, les mouvements du cœur sont faibles, de sorte que ce sont les mouvements de l'artère coronaire matérialisée par l'ensemble guide-cathéter qui sont estimés (champ tracé en bleu). Le mouvement du diaphragme est anatomiquement inexact à cette date également.

À l'inverse, pour la troisième image présentée, se référant à un instant où le diaphragme apparaît statique et le cœur en mouvement, le mouvement de faible amplitude de ce dernier est bien appréhendé (champ tracé en bleu), tout comme celui du fond statique (champ tracé en rouge) et de l'ensemble cathéter-guide (champ tracé en vert).

Après l'injection du produit de contraste, l'artère coronaire est très contrastée et son mouvement devient dominant. Les modèles de mouvement estimés sont pertinents quand les mouvements sont approximativement constants sur deux intervalles de temps successifs.

L'analyse de ce dernier exemple montre donc que la méthode d'estimation du mouvement en situation de bi-transparence distribuée réagit tout aussi bien à la présence d'outils interventionnels et d'injections que celle en situation de bi-transparence simple.

Globalement, les mouvements estimés sont cohérents avec l'observation dans 80% des cas environ avant injection (ce qui est très satisfaisant étant donné que le mouvement des outils interventionnels est souvent pris en compte), et dans 50 à 60% des cas après injection.

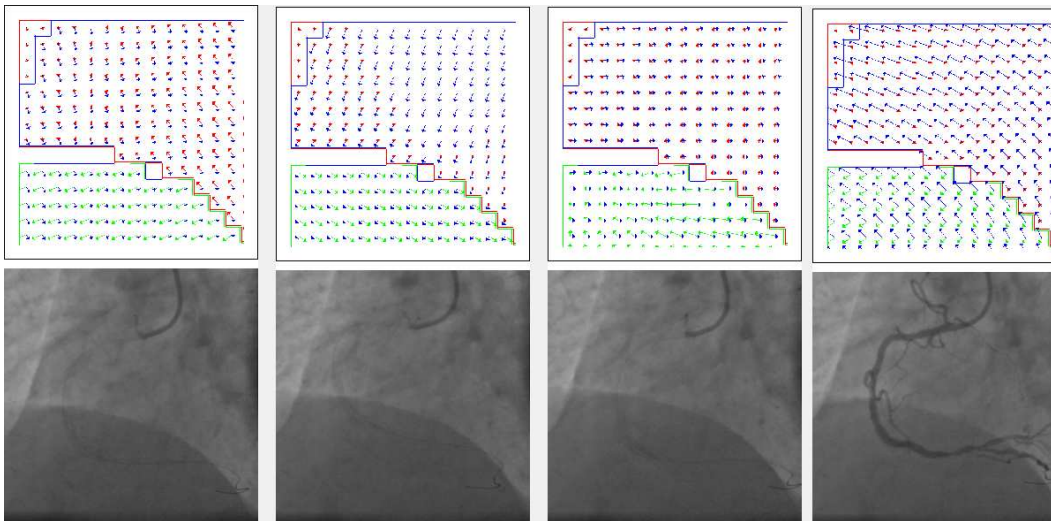


FIG. 4.13: Estimation du mouvement de la séquence diagnostique 33, en situation de bi-transparence distribuée (voir texte principal).

4.4 Conclusion

Nous avons traité dans ce chapitre le problème de l'estimation de mouvements transparents sur des images contenant un nombre quelconque de couches. Nous nous sommes basés sur des observations de séquences réelles révélant que les transparences triples locales mettant en jeu trois couches significatives étaient rares, pour introduire le concept de *bi-transparence distribuée*. Il s'agit des configurations qui peuvent être segmentées en régions mono- ou bi-transparentes.

Ce chapitre s'est concentré sur l'estimation des mouvements en bi-transparence distribuée avec une segmentation en couches donnée par l'utilisateur. La méthode définie commence par calculer les deux modèles de mouvement présents dans chaque région bi-transparente (par mise en correspondance par blocs puis transformée de Hough), puis évalue un modèle de mouvement affine pour chaque couche (au terme d'un mécanisme alterné d'attribution des modèles de mouvement estimés par région aux couches correspondantes, et de calcul d'un modèle affine pour chaque couche). Enfin, une estimation paramétrique globale robuste complète le procédé au niveau de l'image entière.

Nous avons validé cette méthode sur des séquences synthétiques réalistes, proches de la réalité clinique en ce qu'elles mettent en jeu des zones de transition entre couches adjacentes. Les résultats sont satisfaisants pour des images diagnostiques et intéressants pour des images interventionnelles, quand on tient compte du fait que le schéma de génération aléatoire des mouvements favorise les situations délicates où les mouvements sont proches. Finalement, quatre séquences cliniques ont été traitées sur lesquelles les estimations de mouvement sont satisfaisantes et pour lesquelles des structurations par le mouvement apparaissent plus pertinentes qu'une segmentation sur des critères purement anatomiques. Cela dit, ces résultats supposent une intervention de l'utilisateur, ce qui n'est pas compatible avec un certain nombre d'applications dont le débruitage. Nous allons donc présenter dans le chapitre 5 un schéma joint d'estimation et de segmentation par le mouvement en situation de bi-transparence distribuée.

Chapitre 5

Estimation jointe du mouvement et de la segmentation en couches en situation de bi-transparence distribuée

La méthode présentée au chapitre précédent permet des estimations fiables de mouvements sur des séquences d'images cliniques avec transparence dans le cas général. Néanmoins, elle nécessite l'intervention d'un opérateur qui lui fournisse la segmentation de l'image en ses différentes couches. Si une telle interaction est envisageable pour certaines applications, il nous faut définir une méthode capable de traiter les images de manière complètement automatique pour une utilisation plus large, à des fins de débruitage notamment. Ce chapitre présente une méthode répondant à cet objectif.

Nous introduisons dans la partie 5.1 le formalisme permettant de manière jointe une estimation des modèles de mouvement et une segmentation de l'image en couches. Nous précisons l'étape d'initialisation choisie en 5.2. Nous détaillons ensuite en 5.3 comment nous déterminons et faisons évoluer le nombre de couches présentes dans l'image. Nous étudions en 5.4 le comportement de l'algorithme sur des régions mono-couches et des régions multicouches comportant plus de deux couches. Nous concluons ce chapitre par des résultats expérimentaux à la fois sur des données synthétiques et des exemples réels.

5.1 Un problème joint d'estimation et de segmentation par le mouvement

5.1.1 Approche

Nous cherchons à segmenter l'image en ses différentes couches transparentes. Une segmentation classique s'appuyant sur les niveaux de gris de l'image semble très délicate étant donné le niveau de bruit très important des séquences considérées et leur faible contraste. Seule une approche intégrant beaucoup d'*a priori* sur l'anatomie du patient serait envisageable, mais elle serait lourde à mettre en place, et nécessiterait certainement un dialogue avec l'utilisateur afin d'indiquer l'anatomie imagée.

En outre, elle ne nous fournirait pas exactement le type de segmentation que nous recherchons puisque nous ne nous intéressons pas spécifiquement aux organes mais aux *couches* de mouvement cohérent. Nous pouvons d'ailleurs préciser la définition d'une couche dans le cadre de notre approche paramétrique :

Définition 3 *Nous considérons qu'une **couche** est un ensemble physique animé d'un mouvement **affine** cohérent, sous l'angulation d'acquisition.*

Il semble donc naturel d'opter pour une segmentation *relative au mouvement* dans l'image. Rappelons l'approche de Toro et al. [Toro 00b] pour résoudre ce problème, décrite au chapitre 2. Adaptant une méthode proche de celle développée par [Bergen 92a] aux images transparentes, les auteurs procèdent en deux étapes :

- estimation d'un champ dense de déplacements transparents,
- regroupement des vecteurs en différents champs de mouvement affines.

Même si la mise en œuvre de ces deux étapes est moyennement convaincante (utilisation de la méthode du simplex localement pour estimer le champ dense double des déplacements, puis de la technique EM pour regrouper les vecteurs de déplacement en champs affines), il s'agit de la seule segmentation explicite en couches à notre connaissance.

Nous allons également réaliser une segmentation de l'image par le mouvement, un modèle de mouvement affine étant associé à chaque couche transparente. Par contre, plutôt que de procéder séquentiellement comme dans [Toro 00b] (estimation puis segmentation), nous préférons proposer un schéma de segmentation et d'estimation *jointes*. En effet, il va de soit que l'estimation des mouvements transparents sera plus fiable si la segmentation en couches est meilleure, et réciproquement. Les deux tâches sont donc intrinséquement liées, et une approche jointe paraît plus stable et plus efficace.

Ce problème joint d'estimation et de segmentation par le mouvement a déjà été largement étudié [Murray 87, Bouthemy 93, Odobez 98, Paragios 05] dans le cas de séquences d'images vidéo. Il n'a par contre à notre connaissance jamais été traité pour les séquences d'images transparentes.

D'autre part, nous ne cherchons pas à segmenter l'image pixel à pixel, mais bloc par bloc. Nous voulons en effet que l'algorithme ait un coût calculatoire raisonnable pour qu'une adaptation temps réel ne soit pas irréaliste. Puisque l'image est très bruitée et peu contrastée, nous pensons de plus qu'une segmentation directe au pixel près serait très instable. Il serait par contre envisageable, pour des applications ayant des contraintes de temps de calcul moins fortes, de réduire progressivement la taille des blocs, pour atteindre le niveau des pixels si besoin est.

Nous formalisons ce schéma d'estimation et de segmentation jointes à l'aide des champs de Markov.

5.1.2 Formalisation markovienne

L'image étant divisée en S blocs (de taille typiquement 32×32 pixels), nous cherchons à les étiqueter suivant la paire de couches qui y est présente. Notons $e = \{e(s), s \in S\}$ le champ des étiquettes où s note le site (ici, le bloc), avec $e(s) = (e_1(s), e_2(s))$ représentant le couple des numéros des deux couches présentes au site s . Nous supposons que l'image contient un total de K couches transparentes. À chacune d'entre elles est associé un modèle de mouvement affine de paramètres θ_k , et nous notons $\Theta = \{\theta_k, k = 1, \dots, K\}$.

Nous considérons une structure de voisinage à 4 voisins. L'ensemble C des cliques contient des cliques à deux sites. Les *observations* sont les niveaux de gris de l'image. Le nombre de couches K est bien entendu inconnu. La façon de le déterminer est décrite à la section 5.3.

Cette modélisation est schématisée sur la Fig.5.1.

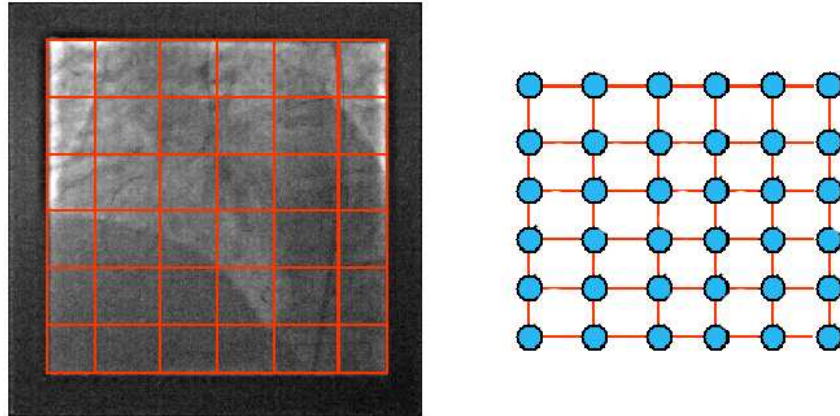


FIG. 5.1: Formalisation du problème par champs de Markov. À gauche une image traitée, divisée en blocs (36 blocs pour les besoins de la clarté du schéma. En fait, l'image étant de taille 1024×1024 , il y aurait véritablement 1024 blocs de taille 32×32 , ce qui autoriserait une segmentation assez précise proche de celle de la Fig.4.2). À droite le modèle de Markov introduit. Les sites sont représentés en bleu, et leurs relations de voisinage sont tracées en orange.

Nous définissons la fonctionnelle d'énergie globale comme suit :

$$\begin{aligned}
 F(e, \Theta) = & \sum_{s \in S} \sum_{\mathbf{p} \in s} \rho(r(\mathbf{p}, \theta_{e_1(s)}, \theta_{e_2(s)})) \\
 & + \mu \sum_{\langle s, t \rangle \in C} \left((1 - \delta(e_1(s), e_1(t))) (1 - \delta(e_1(s), e_2(t))) \right. \\
 & \left. + (1 - \delta(e_2(s), e_1(t))) (1 - \delta(e_2(s), e_2(t))) \right)
 \end{aligned} \tag{5.1}$$

où le symbole de Kronecker $\delta(.,.)$ vaut 1 si les deux arguments sont égaux, et 0 sinon. Elle se décompose en un terme d'attache aux données (sur la première ligne) et un terme de lissage. Le terme d'attache aux données n'est autre que (4.4), introduit dans le chapitre 4. Il impose que l'équation fondamentale de la transparence soit vérifiée au niveau de chaque bloc (avec une tolérance aux outliers venant de l'estimateur robuste ρ , en l'occurrence la fonction de Tukey).

Le second terme impose à la segmentation d'être raisonnablement lisse. Il est constitué de quatre facteurs qui permettent de tester tous les appariements possibles entre les couches de blocs adjacents. Le paramètre μ pondère l'importance relative des deux termes. Autrement dit, on introduit une pénalité μ lorsqu'on crée une frontière entre les sites s et t impliquant

le changement d'une couche seulement, et une pénalité 2μ quand les deux couches sont différentes.

La valeur donnée à μ doit dépendre des images traitées, sans quoi :

- pour des valeurs de μ importantes, il favorisera des champs très lisses sur des images peu contrastées,
- et inversement pour des valeurs de μ faibles, la régularisation aura trop peu d'influence sur des images plus texturées.

C'est pourquoi nous définissons μ comme suit :

$$\mu = \lambda \operatorname{med}_{s \in S} \sum_{\mathbf{p} \in s} \rho(r(\mathbf{p}, \hat{\theta}_{e_1(s)}, \hat{\theta}_{e_2(s)})). \quad (5.2)$$

Nous pouvons alors régler λ en fonction du type d'application projetée, indépendamment des caractéristiques de la séquence d'images particulière considérée. Un λ faible favorisera des estimations du mouvement guidées par les données, tandis qu'une valeur plus importante mènera à des segmentations plus lisses.

Il est d'ailleurs envisageable de faire évoluer λ au fil de l'algorithme. Les estimations des champs de mouvements devenant plus fiables, nous pouvons choisir de donner plus de poids au terme de lissage (en augmentant la valeur de λ) pour assurer une segmentation plus régulière. Nous avons testé différents réglages possibles de λ en section 5.5 sur les images synthétiques, pour finalement choisir des valeurs de l'ordre de 0,5.

D'autre part, évaluer (5.2) est peu coûteux en temps de calcul puisque les résidus $\rho(r(\mathbf{p}, \hat{\theta}_{e_1(s)}, \hat{\theta}_{e_2(s)}))$ sont déjà calculés par ailleurs, et que le médian est ensuite calculé au niveau des blocs, et non au niveau des pixels de l'image.

5.1.3 Minimisation de l'énergie globale

L'énergie (5.1) se minimise itérativement. Nous procédons alternativement comme suit :

- À étiquettes fixées, nous déterminons les paramètres des modèles affines de mouvement minimisant (5.1).
- À paramètres de mouvement fixés, nous déterminons l'étiquetage optimal.

Ce schéma est itéré plusieurs fois. Le critère d'arrêt retenu, pour limiter le volume de calcul, est un nombre d'itérations maximal n imposé. Différentes valeurs de n sont testées dans la section 5.5, et nous établissons que la valeur $n = 5$ offre un bon compromis.

Estimation des modèles affines à segmentation donnée. Elle a fait l'objet du chapitre 4. De fait, quand les étiquettes sont fixées, le terme de lissage de (5.1) est constant, et il suffit de minimiser le terme d'attache aux données avec la méthode présentée au paragraphe 4.2.

Détermination des étiquettes à paramètres affines fixés. Lorsque les modèles affines de mouvement sont fixés, (5.1) devient une fonction du champ des étiquettes e seul. Comme il n'est pas possible de tester tous les scénarios possibles (il y en a K^{2m} si m est le nombre de blocs considérés), même avec un petit nombre de sites, nous minimisons l'énergie globale $F(e)$ par la méthode *ICM* (Iterative Conditional Modes, modes conditionnels itérés).

La méthode retenue consiste à visiter aléatoirement les sites, pour chacun d'entre eux de

tester les K^2 étiquetages possibles¹, et de garder celui qui donne localement l'énergie la plus faible. L'algorithme s'arrête lorsqu'il n'y a plus aucun changement d'étiquette, ou bien quand un nombre maximal d'itérations a été effectué.

Une telle technique de minimisation permet d'atteindre un minimum local qui dépendra de l'initialisation choisie.

L'implantation de l'ICM est relativement peu coûteuse en temps de calcul. Il faut au préalable calculer les images interpolées suivant les $K^2/2$ configurations possibles (une dizaine en pratique pour $I(., t)$ et $I(., t - 1)$). Chaque itération de l'ICM revient alors à calculer le résidu r sur un bloc pour tous les étiquetages possibles, ce qui est excessivement rapide.

Nous résumons le schéma joint d'estimation de mouvements paramétriques et de segmentation en couches à l'algorithme 3.

Algorithme 3 Algorithme du schéma joint d'estimation de mouvements paramétriques transparents et de segmentation en couches.

Itérativement,

- **Estimation des modèles affines de mouvement à étiquetage fixé**
 - Minimisation de l'énergie globale par les moindres carrés pondérés itérés.
 - Utilisation de pyramides de B-splines.
 - Résolution incrémentale de type Gauss-Newton à partir d'une linéarisation du résidu au second ordre.
 - Recherche d'une direction de descente par le gradient conjugué.
 - Calcul du pas exact en annulant un polynôme de degré 4.

- **Détermination du champ d'étiquettes (segmentation) à paramètres de mouvement affines fixés**

Par la technique ICM.

- Visite aléatoire de sites (blocs).
- En chaque site, test de l'ensemble des étiquetages possibles.
- En chaque site, test de l'ensemble des étiquetages possibles.

Nous expliquons dans les sous-sections suivantes

- comment nous initialisons le nombre de couches K , leur modèle de mouvement et la segmentation de l'image en couches ;
- comment nous faisons évoluer K ;
- et comment cet algorithme réagit sur des blocs à une seule couche ou à plus de deux couches.

¹En fait, pour des raisons de symétrie, il n'y en a que $K^2/2$ à considérer.

5.2 Initialisation

Il nous faut initialiser le schéma itératif d'estimation et de segmentation jointes de manière aussi adéquate que possible pour assurer sa convergence vers un minimum local "pertinent". À la différence du chapitre 3, nous ne cherchons pas seulement des valeurs initiales pour les modèles de mouvement des couches, mais également une première segmentation en couches, et une estimation du nombre de couches en présence.

Nous reprenons l'idée d'exploiter les résultats de la mise en correspondance par blocs adaptée à la transparence. Elle s'applique en effet sur un voisinage assez localisé pour que l'on puisse considérer qu'une large majorité des blocs se trouve en situation de bi-transparence. Il faut pour cela bien sûr ignorer les blocs situés à cheval sur la frontière entre deux couches (qui sont largement minoritaires) et les rares blocs pouvant recouvrir une configuration à 3 couches ou plus.

À l'issue de la mise en correspondance par blocs, nous disposons d'un ensemble de vecteurs de déplacement sans structure, et potentiellement entâché d'outliers (dus au bruit sur des zones peu texturées, et également aux blocs qui n'étaient pas en situation bi-couche), comme illustré à gauche de la figure 5.2. Nous souhaitons déterminer à partir de cet ensemble de vecteurs de déplacement :

- le nombre de couches présentes dans l'image ;
- une première approximation des modèles de mouvement paramétriques des couches ;
- une segmentation de l'image en couches.

Nous choisissons d'utiliser à nouveau la transformée de Hough, dans un espace affine simplifié (deux composantes de translation et un terme de divergence). Nous faisons à nouveau l'hypothèse que ce modèle permet d'approcher raisonnablement les mouvements anatomiques réels à ce stade d'initialisation.

Dans l'espace de Hough considéré, chaque déplacement vote pour un ensemble de paramètres permettant de l'expliquer (définissant une droite). Au lieu d'incrémenter la matrice d'accumulation de manière entière, nous associons à chacun de ces déplacements un coefficient compris entre 0 et 1 traduisant la confiance sur le déplacement, comme expliqué à la section 3.4.

Les intersections principales des différentes droites fournissent les paramètres des modèles de mouvement expliquant un nombre significatif de déplacements calculés par la mise en correspondance par blocs. Alors que dans le chapitre 3, nous retenions les deux intersections les plus significatives pour extraire les modèles de mouvement des deux couches, nous choisissons d'extraire sur les images en *bi-transparence distribuée* itérativement tous les modèles correspondant aux intersections jugées représentatives.

Ceci se traduit donc par une extraction itérative des intersections les plus significatives (à deux niveaux de résolution) comme précisé ci-après. Nous introduisons pour chacune d'elles une nouvelle couche, et attribuons son étiquette à chacun des blocs dont les déplacements sont à moins de 2 pixels du modèle estimé. L'intersection est considérée comme représentative si au moins cinq déplacements ont contribué à la valeur de la cellule correspondante dans la matrice d'accumulation. L'extraction itérative est stoppée dès qu'une couche non représentative est détectée.

Nous illustrons les résultats de la mise en correspondance par blocs et l'extraction de modèles paramétriques à la Fig.5.2.

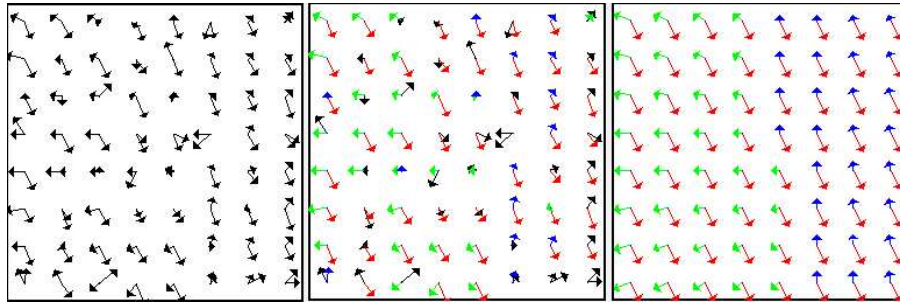


FIG. 5.2: À gauche les déplacements issus de la mise en correspondance par blocs avec transparence bi-couche. Au centre, les champs correspondant aux modèles affines du mouvement extraits par la transformée de Hough et les vecteurs sur lesquels ils s'appuient. À droite, les vrais déplacements.

Cette étape permet d'estimer le nombre de couches en présence (qui sera mis à jour comme expliqué en 5.3), et d'initialiser leurs paramètres de mouvement. Il reste encore à initialiser la segmentation de l'image. Tous les vecteurs de déplacement n'ont en effet probablement pas été attribués à l'une des couches à ce stade (et normalement pas les outliers!).

Nous attribuons à chaque bloc l'étiquette se traduisant par le résidu le plus petit au sens de (4.3). En d'autres termes, nous minimisons l'énergie (5.1) sans son terme de lissage, ce qui revient à considérer un critère de maximum de vraisemblance.

L'algorithme 4 résume le schéma d'initialisation.

Algorithme 4 Algorithme de l'initialisation de l'estimation de mouvement transparent en situation de bi-transparence distribuée sans *a-priori*.

- Mise en correspondance par blocs avec transparence bi-couche.

- Extraction des modèles affines de mouvement simplifiés par transformée de Hough.
 1. Attribution d'un coefficient de confiance à chaque déplacement issu de la mise en correspondance par blocs. Puis, itérativement,
 2. Extraction (multirésolution) du modèle affine le plus représenté,
 3. Nouvel étiquetage des blocs dont un des deux déplacements est situé à moins de 2 pixels du champ extrait,
 4. Répétition si au moins 5 blocs ont été attribués à la couche introduite.

- Initialisation de la segmentation en couche.

Par maximum de vraisemblance (minimisation du terme d'attache aux données de (5.1))

5.3 Gestion du nombre de couches

Le schéma markovien que nous avons défini suppose connu le nombre total de couches K présentes dans l'image. Or, si nous en connaissons une première évaluation grâce à la transformée de Hough utilisée dans l'étape d'initialisation, il est nécessaire de faire évoluer au besoin cette valeur au cours de la minimisation de la fonction $F(e, \Theta)$.

Pour ce faire, nous introduisons un mécanisme permettant de réduire ou d'augmenter le nombre de couches. Nous présentons ces deux alternatives dans les sous-sections suivantes.

5.3.1 Mécanisme de réduction du nombre de couches

Nous pouvons avoir deux raisons de diminuer le nombre de couches : une couche ne recouvre aucune réalité physique, ou elle n'est que le doublon d'une autre couche et induit donc une sur-segmentation.

Élimination des couches sans réalité physique. Dans ce cas, cette couche a été introduite au "hasard des regroupements" de déplacements sans doute erronés issus de la mise en correspondance par blocs. Elle ne présente en général aucune régularité spatiale. La couche en question est donc généralement fragmentée, et l'évaluation du terme d'attache aux données lui correspondant fournit une valeur peu significative. Elle sera nécessairement fortement pénalisée par le terme de lissage de (5.1). Il est probable le nombre de blocs qui lui sont affectés diminue nettement dès la première itération de minimisation par ICM. Nous décidons en pratique d'éliminer toute couche contenant moins de 5 blocs au terme d'une itération complète de l'ICM.

Fusion des couches sursegmentées. Lorsque deux couches sont détectées là où seule une est effectivement présente, les deux modèles de mouvement estimés deviennent très proches. Nous choisissons donc de fusionner deux couches quand les champs de déplacements issus de leurs modèles de mouvement respectifs diffèrent en moyenne de moins de 2 pixels sur l'union des blocs qu'elles contiennent.

Ce seuil est choisi relativement élevé car nous recherchons dans la pratique des mouvements assez clairement différents.

5.3.2 Utilisation pratique de l'ajout de couche

Si les mécanismes de réduction du nombre de couches décrits plus haut sont faciles à mettre en œuvre, une procédure plus complexe pour décider de l'ajout d'une couche doit être envisagée. Nous avons développé une technique qui détecte les blocs mal étiquetés sur la base des poids issus de l'estimation robuste des modèles affines de mouvement, et qui décide de l'ajout d'une nouvelle couche quand un certain nombre de ces blocs a été repéré. Ce mécanisme d'ajout des couches est présenté en détail en annexe C.

Cependant, pour estimer de manière fiable les mouvements de la nouvelle couche introduite, il faut itérer la minimisation de (5.1), ce qui double quasiment le temps de calcul de la méthode d'estimation de mouvement dans son ensemble. Nous avons donc cherché à rendre son utilisation aussi peu fréquente que possible.

Ainsi, nous n'effectuons ce test qu'une fois seulement, à la fin de l'exécution normale

du schéma markovien. Ceci permet de se baser sur une segmentation et des modèles de mouvement fiables, qui limitent le risque d'ajouter une couche à mauvais escient.

D'autre part, nous avons fait en sorte d'être le plus souvent possible en situation de sur-segmentation lors de l'initialisation, et le moins souvent possible en situation de sous-segmentation. Outre le fait que le premier cas est infiniment plus facile à traiter que le second, les conséquences pratiques d'une sur-segmentation pour l'application finale sont souvent moins graves puisqu'elle n'empêche pas d'estimer correctement les mouvements.

Ceci explique les réglages choisis de la transformée de Hough², qui mènent à de fréquentes sur-segmentations, mais permettent de n'oublier que très rarement des couches vraiment présentes dans l'image.

Nous avons constaté qu'avec ce type de réglages, le mécanisme d'ajout de couches n'améliorait quasiment pas la précision d'estimation. Les résultats rapportés en annexe C mettent en évidence dans le meilleur des cas un gain infime de précision au prix d'une augmentation conséquente du temps de calcul.

En conséquence, les deux modules qui nous permettent d'estimer le nombre de couches présentes dans la séquence d'images sont les suivants :

- **Initialisation.** Le nombre de couches présentes dans la séquence d'images est initialisé par le nombre de modèles de mouvement extraits par la transformée de Hough appliquée aux vecteurs de déplacement issus de la mise en correspondance par blocs. Elle est réglée pour favoriser les sur-segmentations.
- **Réduction du nombre de couches.** En cours d'algorithme, les couches qui étiquettent moins de 5 blocs sont éliminées et les blocs correspondant réétiquetés à partir des modèles de mouvement restants. Lorsque les champs associés à deux modèles de mouvement sont distincts de moins de 2 pixels en moyenne, les couches correspondantes sont fusionnées.

Nous rappelons à la Fig.5.3 les différentes étapes de notre méthode d'estimation de mouvement en situation de bi-transparence distribuée. Elle est appliquée à l'une des images réalistes en situation de bi-transparence distribuée de la section 4.3.1.1. Nous ne nous intéressons pas à la zone de transition.

Nous pouvons y noter que les déplacements calculés par la mise en correspondance par blocs sont très bruités. La confiance associée à chacun d'entre eux est représentée par le niveau de gris du bloc correspondant. Nous remarquons que la plupart des déplacements auxquels sont associés une haute confiance rendent compte correctement des mouvements réels.

La détermination des modèles de mouvements affines simplifiés permet bien de trouver les trois mouvements en présence. La translation (en rouge sur la vérité terrain) donne lieu à deux modèles proches (en bleu et en rouge sur la figure présentant les modèles de mouvement déterminés par la transformée de Hough), tandis que les deux autres mouvements (affines) sont captés en partie (par les modèles en vert et jaune). Par contre, le modèle violet ne correspond à aucune couche physique. L'attribution de ces modèles aux différents blocs de l'image mène à une segmentation encore très confuse.

La première estimation des modèles affines complets à segmentation donnée a peu d'im-

²Pas de discrétisation relativement grossier, et extraction de paramètres itérée dès que 5 blocs au moins ont été associés à la couche dernièrement extraite.

pact puisque la segmentation est mauvaise. Le premier réétiquetage avec lissage permet de commencer à regrouper les blocs en couches de manière cohérente. L'itération suivante de la minimisation de l'énergie globale voit la disparition du modèle de mouvement de la couche "violette" alors que le modèle tracé en bleu s'approche du mouvement affine réel de la couche présente en bas de l'image. Les trois itérations suivantes permettent d'améliorer la segmentation de l'image en ses différentes couches, et avec elle l'estimation des paramètres des modèles de mouvement.

Le résultat final est excellent puisqu'il fournit les mouvements présents en chaque bloc. La seule erreur est une sur-segmentation de la couche inférieure, mais cela a peu d'impact sur toutes les applications qui s'intéressent aux mouvements plus qu'à la segmentation en couches proprement dite.

5.4 Comportement sur les blocs à une seule couche et à plus de deux couches

L'algorithme décrit jusque-là permet de segmenter l'image en ses différentes couches, et d'estimer leur mouvement, *sous l'hypothèse de la présence de deux couches dans chaque bloc*. Nous précisons dans cette section quel est son comportement sur des blocs mono-couches d'une part et à plus de deux couches d'autre part (pour simplifier, nous parlerons de blocs *multi-couches* pour ces derniers). Nous expliquerons ensuite pourquoi, alors que des moyens de détecter les blocs dans ces deux configurations peuvent être imaginés (Annexe D), nous nous contentons finalement de l'algorithme d'estimation de mouvements en bi-transparence distribuée tel qu'il est.

5.4.1 Comportement sur un bloc mono-couche

Rappelons l'expression du résidu (2.39) exploité dans l'estimation des mouvements bi-transparents :

$$\begin{aligned} r(\mathbf{p}, \mathbf{w}_1, \mathbf{w}_2) &= I(\mathbf{p}, t+1) - I(\mathbf{p} + \mathbf{w}_1, t) \\ &\quad - I(\mathbf{p} + \mathbf{w}_2, t) + I(\mathbf{p} + \mathbf{w}_2 + \mathbf{w}_1, t-1). \end{aligned} \quad (5.3)$$

Il se réduit dans le cas d'une seule couche à :

$$\begin{aligned} r(\mathbf{p}, \mathbf{w}_1, \mathbf{w}_2) &= I_1(\mathbf{p}, t+1) - I_1(\mathbf{p} + \mathbf{w}_1, t) \\ &\quad - I_1(\mathbf{p} + \mathbf{w}_2, t) + I_1(\mathbf{p} + \mathbf{w}_2 + \mathbf{w}_1, t-1). \end{aligned} \quad (5.4)$$

Supposons que le mouvement de la couche unique présente dans le bloc soit capté par un des deux modèles de mouvement, par exemple celui induisant \mathbf{w}_1 . Nous reconnaissons dans (5.4) la conservation *classique* de l'intensité (mono-couche) pour le pixel \mathbf{p} entre les dates $t+1$ et t , et pour le pixel $\mathbf{p} + \mathbf{w}_2$ entre les dates t et $t-1$.

En d'autres termes, si \mathbf{w}_1 est la vitesse de la couche unique au point considéré, le résidu est censé être nul *quelle que soit la deuxième vitesse \mathbf{w}_2* ³ :

$$r(\mathbf{p}, \mathbf{w}_1, \mathbf{w}_2) = 0, \forall \mathbf{w}_2. \quad (5.5)$$

³Du moins tant que la vitesse de la couche unique vaut aussi \mathbf{w}_1 en $\mathbf{p} + \mathbf{w}_2$, ce qui suppose une translation locale à l'échelle de \mathbf{w}_2 . Nous retrouvons l'hypothèse nécessaire à l'établissement de l'équation de contrainte du mouvement transparent (2.39).

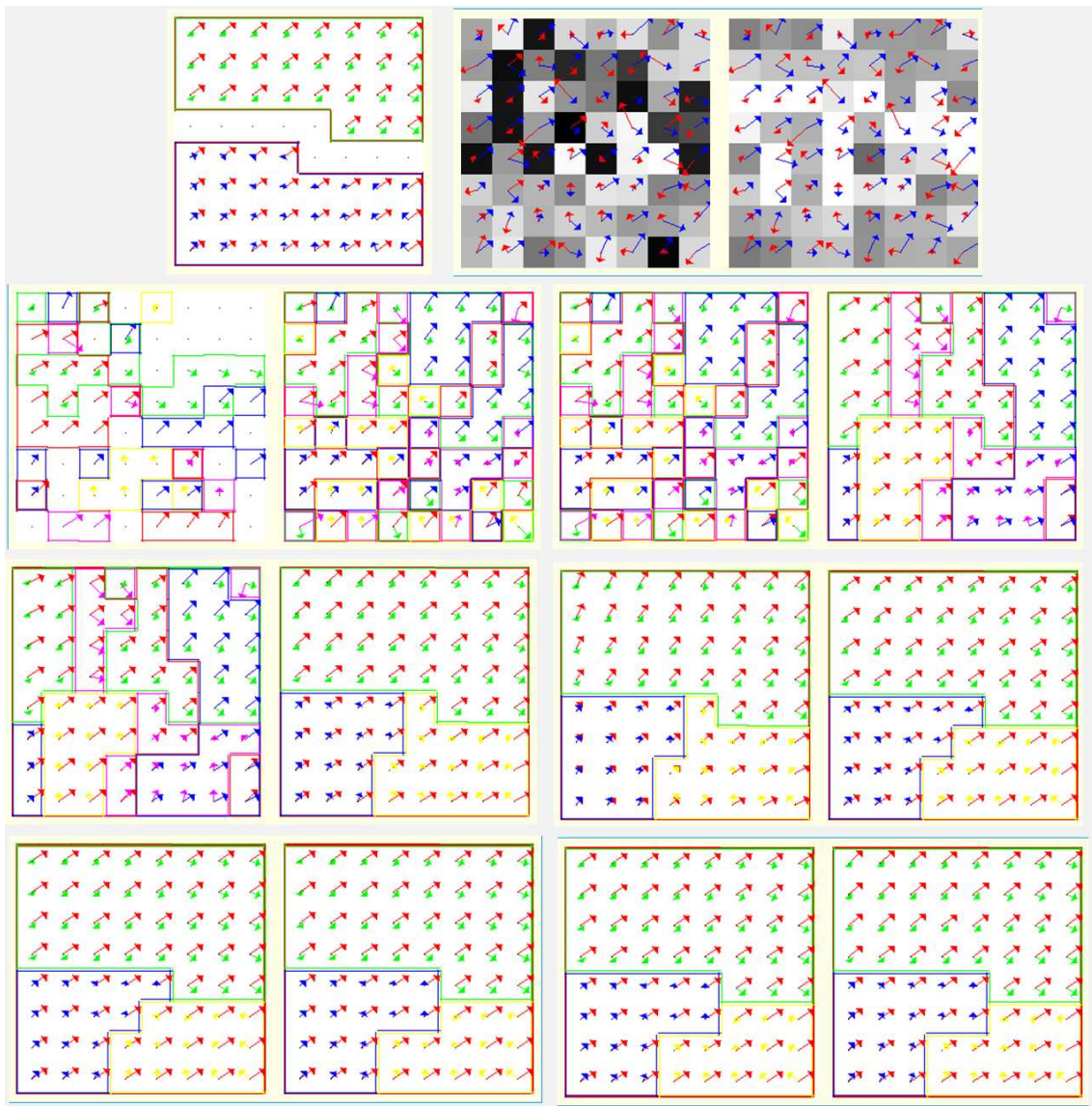


FIG. 5.3: Étapes de l'algorithme d'estimation-segmentation de mouvement en bi-transparence distribuée. *Ligne du haut de gauche à droite* : la vérité-terrain (champs de vitesses utilisés pour générer la séquence réaliste d'images, comme expliqué dans le paragraphe 4.3.1.1, comprenant une zone mono-couche de transition à laquelle nous ne nous intéressons pas ici) ; les déplacements issus de la mise en correspondance par blocs. Le niveau de gris des blocs montre le coefficient de confiance associé aux vecteurs en bleu (au centre) et en rouge (à droite). Notons que les couleurs utilisées ne servent ici qu'à visualiser les résultats puisqu'aucun étiquetage n'a encore eu lieu. *Deuxième ligne, de gauche à droite* : modèles affines simplifiés extraits par la transformée de Hough et attribués aux blocs qui ont voté pour eux, modèles affines simplifiés attribués aux blocs dont ils minimisent le résidu, puis première itération de la minimisation de l'énergie (5.1) avec tout d'abord l'estimation des modèles affines complets à segmentation donnée, puis le réétiquetage à modèles de mouvement donnés. *Troisième ligne* : Itérations 2 et 3 de la minimisation de l'énergie globale (5.1). *Dernière ligne* : Itérations 4 et 5 de la minimisation de l'énergie globale (5.1).

Par conséquent :

- le schéma d'estimation proposé permet de calculer les mouvements de la couche présente sur les blocs mono-couches.
- cependant, la méthode va également estimer un modèle fictif, sans signification physique sur ce bloc, dont l'apparition peut avoir deux conséquences gênantes :
 - elle peut entraîner une segmentation inexacte, ce qui est problématique pour certaines applications ;
 - cette segmentation peut gêner l'estimation du mouvement, par exemple en complétant et validant des couches peu représentatives, ce qui limitera l'effet du lissage introduit dans l'énergie (5.1). Ce problème est toutefois secondaire.

5.4.2 Comportement sur un bloc multi-couche.

Comme nous l'avons expliqué en début de chapitre, c'est précisément la rareté de la configuration multi-couche qui a fait privilégier la notion de bi-transparence distribuée pour appréhender le cas général. Lorsque nous rencontrons malgré tout un bloc multi-couche, par exemple tri-couche, notre critère est alors défaillant. Le résidu s'écrit dans ce cas-là :

$$\begin{aligned}
 r(\mathbf{p}, \mathbf{w}_1, \mathbf{w}_2) &= I(\mathbf{p}, t + 1) - I(\mathbf{p} + \mathbf{w}_1, t) \\
 &\quad - I(\mathbf{p} + \mathbf{w}_2, t) + I(\mathbf{p} + \mathbf{w}_2 + \mathbf{w}_1, t - 1) \\
 &= I_1(\mathbf{p}, t + 1) - I_1(\mathbf{p} + \mathbf{w}_1, t) \\
 &\quad - I_1(\mathbf{p} + \mathbf{w}_2, t) + I_1(\mathbf{p} + \mathbf{w}_2 + \mathbf{w}_1, t - 1) \\
 &\quad + I_2(\mathbf{p}, t + 1) - I_2(\mathbf{p} + \mathbf{w}_1, t) \\
 &\quad - I_2(\mathbf{p} + \mathbf{w}_2, t) + I_2(\mathbf{p} + \mathbf{w}_2 + \mathbf{w}_1, t - 1) \\
 &\quad + I_3(\mathbf{p}, t + 1) - I_3(\mathbf{p} + \mathbf{w}_1, t) \\
 &\quad - I_3(\mathbf{p} + \mathbf{w}_2, t) + I_3(\mathbf{p} + \mathbf{w}_2 + \mathbf{w}_1, t - 1).
 \end{aligned} \tag{5.6}$$

$$\tag{5.7}$$

- Une situation favorable apparaît lorsqu'une des trois couches, par exemple I_3 , est homogène sur le bloc considéré (relativement aux deux autres). Dans ce cas, les deux dernières lignes de l'équation (5.7) se comportent "comme un bruit additif". Cela permet de considérer que l'on reste approximativement en situation de bi-transparence distribuée, et d'estimer les mouvements des deux autres couches assez correctement. En tout état de cause, la présence de la couche 3 sur le bloc ne sera pas détectée, et la segmentation partiellement faussée. Il s'agit d'ailleurs là d'un problème plus général, qui concerne les régions homogènes, et sur lequel nous allons revenir.
- Dans le cas général, minimiser (5.7) sur un bloc en situation de multi-transparence n'a pas de sens physique. Le résidu est alors important quelles que soient les mouvements estimés, ce qui se traduira par des poids faibles dans le schéma d'estimation robuste. L'information de ce bloc sera alors ignorée, ce qui évitera de fausser l'estimation des modèles de mouvement affines par ailleurs. Par contre, un étiquetage sans grand sens physique risquera d'avoir lieu sur ce bloc. Puisque cette configuration est rare, nous pouvons espérer que les contraintes de lissage seront prédominantes et permettront un étiquetage approprié du bloc selon deux des trois couches en exploitant le contexte spatial de ce bloc.

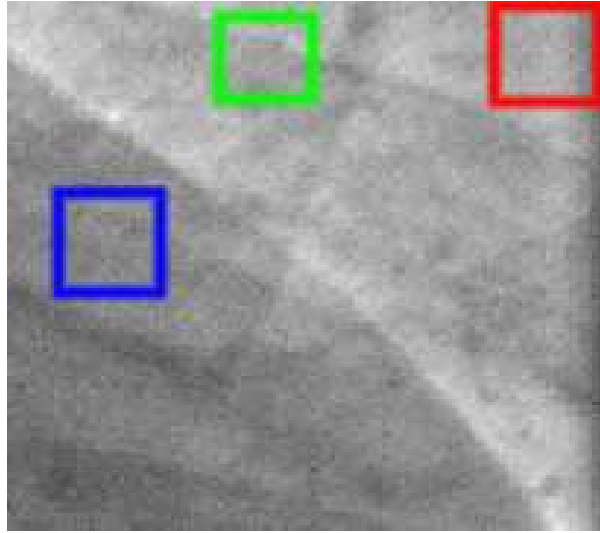


FIG. 5.4: Image clinique avec un bloc en situation monocouche (rouge), un en situation bi-couche visible (vert) et un en situation bi-couche difficilement détectable (bleu).

5.4.3 Pertinence de l'adaptation du schéma d'estimation-segmentation pour détecter les configurations hors bi-couches

Nous proposons deux moyens de détecter les blocs mono- et multi-couches, et expliquons comment adapter le schéma d'estimation-segmentation en conséquence en annexe D. Leur efficacité est montrée sur des images-test de textures bruitées, ainsi que des images vidéos classiques en annexe E.

La méthode de détection des blocs mono-couches cherche à repérer les vitesses fictives (celles qui ne correspondent pas à une couche réelle), en s'appuyant sur le fait que le résidu ne change pratiquement pas si on les remplace par d'autres vitesses choisies aléatoirement. À l'inverse, une vitesse correspondant à une couche réelle donnera un résidu très différent selon que l'on choisisse sa valeur réelle ou des valeurs aléatoires.

Néanmoins, si nous observons effectivement des comportements très différents dans ces deux cas lorsque les deux couches sont assez texturées, ils sont beaucoup plus difficiles à distinguer sur des images cliniques. Si en plus les modèles de mouvement sont approchés de manière imparfaite, le comportement des résidus correspondant à des couches fictives et réelles deviennent très difficiles à distinguer.

Plus intuitivement, la figure 5.4 illustre la difficulté de la tâche de détection de ces configurations. Elle présente une image clinique sur laquelle trois blocs sont dessinés. L'un d'entre eux (le rouge) est en situation mono-couche, correspondant aux poumons. Le bloc vert est en situation bi-couche (poumons et côtes) avec deux couches relativement distinguables. Sa configuration pourrait être détectée. Par contre, le bloc bleu se trouve lui-aussi en situation bi-couche mais la couche du cœur y est si homogène qu'elle ne sera pas détectable indépendamment des autres.

Dans le même ordre d'idée, nous proposons de détecter les blocs en situation de multi-couches par la valeur de leur résidu, qui doit être plus élevée que sur un bloc mono- ou

bi-couche. Là aussi, si cela fonctionne bien sur des images texturées, ce critère n'est pas utilisable de manière fiable sur les images cliniques.

Nous choisissons donc de renoncer à adapter notre schéma d'estimation-segmentation de sorte qu'il puisse explicitement inclure la détection des configurations mono- et multi-couches, et ce d'autant plus que l'algorithme original reste robuste à ces configurations :

- les vitesses des blocs mono-couches sont bien estimées ;
- les blocs multi-couches ne gênent pas directement les estimations des modèles de mouvement des différentes couches.

Soulignons également que l'application principale visée (le débruitage compensé en mouvement des images fluoroscopiques) nous autorise à accepter une carte de segmentation imparfaite pourvu que les mouvements vraiment présents soient suffisamment bien estimés.

5.5 Résultats expérimentaux

Comme dans les chapitres 3 et 4, nous présentons dans un premier temps des résultats sur des séquences synthétiques réalistes. Ceci nous permet de fixer les réglages de la méthode, et d'en mieux comprendre la sensibilité. Nous présentons ensuite des résultats sur des séquences cliniques réelles.

5.5.1 Traitement de séquences synthétiques

5.5.1.1 Choix des métriques

Les séquences d'images synthétiques en situation de bi-transparence distribuée sont générées de la manière détaillée en 4.3.1.1. Puisque nous cherchons à mesurer les résultats d'un schéma joint d'estimation de mouvement et de segmentation en couches, la question de la métrique à utiliser pour juger de la qualité de ses résultats se pose. Nous en proposons trois :

- *L'erreur quadratique moyenne (EQM)*. Il s'agit de l'erreur moyennée sur toute l'image entre les vecteurs de mouvement calculés (déduits des modèles de mouvement estimés et dépendant de la segmentation produite) et ceux de la vérité-terrain :

$$EQM = \sqrt{\frac{1}{2N} \sum_{\mathbf{p} \in \mathfrak{S}-T} \|\mathbf{w}_{e_1(\mathbf{p})}(\mathbf{p}) - \mathbf{w}_1(\mathbf{p})\|^2 + \|\mathbf{w}_{e_2(\mathbf{p})}(\mathbf{p}) - \mathbf{w}_2(\mathbf{p})\|^2} \quad (5.8)$$

où T désigne les zones de transition, et N le nombre de pixels de l'image privée des zones de transition. Les vecteurs $\mathbf{w}_1(\mathbf{p})$ et $\mathbf{w}_2(\mathbf{p})$ correspondent aux vitesses de la vérité-terrain. Elles ont été associées à la vitesse estimée qui leur est la plus proche. L'EQM dépend à la fois de la qualité des modèles de mouvement estimés et de la justesse de la segmentation obtenue. Elle a donc l'inconvénient de faire jouer un rôle important à des erreurs de segmentation sur des blocs peu texturés, alors qu'elles sont en fait sans conséquence pour les applications visées comme le débruitage.

- *L'erreur d'estimation du mouvement (EEM)*. Il s'agit de l'erreur moyenne sur l'image entre les déplacements réels et les modèles des mouvements estimés, évalués à partir de la vraie segmentation de l'image (connue par construction).

Si cette métrique ne rend pas compte de la qualité de la segmentation produite (ou

très indirectement, car les modèles de mouvement estimés dépendent de la segmentation obtenue), elle s'avère par contre plus pertinente pour évaluer effectivement les mouvements estimés en tant que tels, ce qui est notre objectif principal.

- *Le résidu final sur la séquence non bruitée (RF)*. Il s'agit de la moyenne quadratique des résidus (2.39) évaluée sur l'ensemble de l'image, en utilisant les modèles de mouvement affines et la segmentation obtenue par la méthode jointe d'estimation-segmentation.

Nous évaluons cet indicateur sur les images non bruitées que nous générons parallèlement aux images réalistes sur lesquelles les estimations de mouvement et la segmentation en couches sont effectuées. Elles leur sont identiques, au bruit près. Cette précaution nous permet d'avoir une métrique très pertinente, car elle mesure l'erreur de prédiction sur l'anatomie par compensation du mouvement, comme détaillé dans le chapitre 7 :

$$RF = \frac{\sqrt{\sum_{\mathbf{p} \in \mathfrak{S}-T} r(\mathbf{p}, \hat{\Theta})^2}}{N}. \tag{5.9}$$

Le RF permet de mesurer si l'estimation des mouvements explique correctement les variations des niveaux de gris des images. Par contre, sa valeur numérique est moins explicite que l'erreur d'estimation sur les vecteurs de mouvement eux-mêmes.

5.5.1.2 Simulations

Réglages de la minimisation de l'énergie globale $F(e, \theta)$. Nos premiers tests ont porté sur le réglage de la minimisation de l'énergie $F(e, \theta)$ (5.1) en ce qui concerne les deux aspects suivants :

- le nombre maximal d'itérations de la boucle formée de l'estimation des modèles de mouvement à étiquetage fixé et de la segmentation à paramètres de mouvement fixés. Nous avons opté pour 3, 5 et 7 itérations globales.
- le facteur λ , permettant de pondérer lissage et attache aux données dans l'énergie globale $F(e, \theta)$ (5.1). Nous avons testé les valeurs 0,3, 0,5 et 0,7, ainsi que des variations en cours de traitement de 0,3 à 0,7 (0,3 à la première itération, 0,4 à la deuxième, puis 0,5, 0,6 et 0,7) et des variations de 0,4 à 0,6 (0,4, 0,45, 0,5, 0,55 et 0,6) si on se limite à 5 itérations.

Les résultats sont rassemblés dans le tableau 5.1 pour 250 générations de séquences d'images. Nous y donnons les valeurs moyennes des trois métriques proposées plus haut (et leur écart-type) sur l'ensemble des séquences de graphies (examen diagnostique, $\sigma = 10$) et des séquences fluoroscopiques (examen interventionnel, $\sigma = 20$).

Nous avons également calculé la valeur du résidu final moyen RF pour les modèles de mouvement de la vérité-terrain (résidu "idéal"). Il vaut 6,69 sur les images diagnostiques, et 6,80 sur les images interventionnelles, ce qui permet de "calibrer" les résultats obtenus⁴. Il y a deux raisons expliquant pourquoi il n'est pas exactement nul, même avec les mouvements exacts :

⁴Notons que sa valeur n'est pas exactement la même sur images graphiques et fluoroscopiques, bien qu'elles soient construites à partir des mêmes "images-couches" et des mêmes mouvements. La simulation de l'acquisition à des niveaux de rayonnement X différents ne se traduit pas seulement par des niveaux de bruit différents, mais aussi par exemple par des niveaux de gris moyen avant encoding différents, ce qui explique que les images idéales "sans bruit" n'aient pas exactement les mêmes propriétés dans ces deux cas.

- L'équation (2.39) sur laquelle est bâtie l'attache aux données de l'énergie $F(e, \theta)$ (5.1) n'est exacte que pour des mouvements localement translationnels, ce qui n'est pas le cas ici pour deux des trois couches. Néanmoins, la violation de cette hypothèse est faible ici (puisque les modèles affines de mouvement utilisés varient de manière douce localement, et puisque les images traitées contiennent de larges zones homogènes) et a donc un impact marginal sur la valeur du résidu "idéal".
- Les interpolations utilisées pour recaler les images se comportent comme des filtres passe-bas. Certaines hautes fréquences de l'image $I(\mathbf{p}, t-1)$ sont atténuées lors de la *génération* de $I(\mathbf{p}, t)$ et $I(\mathbf{p}+1, t)$. De même, le calcul des images $I(\mathbf{p}+\mathbf{u}, t)$, $I(\mathbf{p}+\mathbf{v}, t)$ et $I(\mathbf{p}+\mathbf{u}+\mathbf{v}, t)$ implique l'atténuation de hautes fréquences. Or, comme les images que nous utilisons pour construire les couches proviennent d'examens réels, elles sont entachées de bruit, ce qui se traduit par une texture spatiale à haute fréquence, peu perceptible par un observateur, mais très sensible aux interpolations.

TAB. 5.1: Précision de l'estimation jointe de mouvement et de segmentation en situation de bi-transparence distribuée. Pour comparaison, le RF calculé avec les vitesses réelles vaut 6,69 sur les images de graphie, et 6,80 sur les images fluoroscopiques.

Réglage	Graphie			Fluoroscopie			Temps de calcul (s)
	EQM	EEM	RF	EQM	EEM	RF	
3 it, $\lambda = 0.3$	0,68	0,42	6,67	2,16	1,16	9,35	74
3 it, $\lambda = 0.5$	0,63	0,39	6,64	1,99	1,18	9,06	73
3 it, $\lambda = 0.7$	0,67	0,40	6,68	2,07	1,24	9,30	73
5 it, $\lambda = 0.3$	0,64	0,39	6,67	2,05	1,10	9,18	112
5 it, $\lambda = 0.5$	0,62	0,38	6,61	1,90	1,10	8,86	112
5 it, $\lambda = 0.7$	0,65	0,40	6,69	1,95	1,14	8,84	111
7 it, $\lambda = 0.3$	0,62	0,38	6,60	2,06	1,08	9,00	151
7 it, $\lambda = 0.5$	0,63	0,39	6,61	1,84	1,10	8,79	152
7 it, $\lambda = 0.7$	0,64	0,39	6,64	1,94	1,14	8,85	153
5 it, $\lambda = 0, 2 - 0, 7$	0,61	0,39	6,58	1,98	1,11	8,89	111
5 it, $\lambda = 0, 4 - 0, 6$	0,65	0,40	6,64	1,88	1,10	8,72	111

Nous remarquons qu'il n'y a pas d'écart majeur entre ces différents réglages, quelle que soit la métrique considérée. Le choix de $\lambda = 0.5$ donne les meilleurs résultats. Si les différences sont perceptibles entre les versions limitées à 3 itérations et les autres, nous ne pouvons par contre pas préférer les versions à 7 itérations à celles à 5, tant le gain est minime pour un coût de calcul nettement plus élevé.

Finalement, **nous fixons le nombre maximal d'itérations à 5 et λ à 0,5**. Notons au passage que faire varier λ ne change pas fondamentalement le comportement de la méthode.

Précision d'estimation. Nous présentons les mesures de précision de l'estimation jointe de mouvement et de segmentation pour les trois métriques présentées ci-dessus, pour différents types d'images : graphie ou fluoroscopie, diffusé nul, normal (20%) ou élevé (50%). L'*EQM* révèle une excellente précision sur les images diagnostiques : elle est de l'ordre de 0.60 pixels pour des images diagnostiques avec un niveau de diffusé modéré. L'*EQM*

reste intéressante sur des images interventionnelles (à niveau de diffusé modéré) : elle affiche une précision de l'ordre de 1,9 pixels, ce qui est satisfaisant puisque cette métrique est exigeante en ce qui concerne la précision de segmentation. Ainsi, la dégradation par rapport aux résultats du chapitre 4 est faible (1,9 contre 1,8) alors qu'on ne connaît pas la segmentation⁵. Par contre, comme dans le chapitre précédent, la méthode rencontre ses limites dans le cas le plus difficile ($\sigma = 20$ et diffusé de 50%).

L'EEM, qui met l'accent sur l'estimation des mouvements, est encore meilleure : 0,4 pixels sur les images diagnostiques, et 1,1 sur les images fluoroscopiques. Notons enfin que le résidu moyen est dans le cas des séquences diagnostiques du même ordre que le résidu "idéal", et est dégradé d'environ 20% dans le cas de séquences interventionnelles. Cela dit, cette mesure est difficile à interpréter car on ne sait rien de sa répartition spatiale (régulière ou très localisée).

TAB. 5.2: Précision de l'estimation jointe de mouvement et de segmentation en situation de bi-transparence distribuée, pour différents types d'images (diagnostiques ou interventionnelles, diffusé 0%, 20% ou 50%).

Type d'images	Moyenne				Écart-type		
	EQM	EEM	RF	RF Idéal	EQM	EEM	RF
Graphie Diffusé 0	0,52	0,34	7,70	7,66	0,48	0,24	1,27
Graphie Diffusé 20%	0,62	0,38	6,61	6,55	0,71	0,32	1,02
Graphie Diffusé 50%	0,89	0,56	5,68	5,56	0,70	0,34	1,05
Fluoroscopie Diffusé 0	1,36	0,84	9,46	7,77	0,91	0,49	1,39
Fluoroscopie Diffusé 20%	1,90	1,10	8,86	6,67	1,18	0,65	1,74
Fluoroscopie Diffusé 50%	3,01	1,65	9,00	5,69	1,26	0,90	1,54

Nous présentons également à la figure 5.5 les erreurs commises sur les 250 tests sur des images diagnostiques et fluoroscopiques avec 20% de diffusé, mesurées par l'EQM et l'EEM. Ces diagrammes révèlent que les estimations de mouvement en situation de bi-transparence distribuée sont répétables de manière fiable dans le cas des images diagnostiques. Le nuage de points est par contre plus diffus dans le cas d'images interventionnelles, mais ce phénomène est plus net quand on considère l'EQM que l'EEM. En d'autres termes, si l'estimation de mouvement sur images interventionnelles est répétable de manière fiable, la segmentation associée sera d'une qualité beaucoup plus variable.

⁵L'estimateur à segmentation donnée a toutefois l'avantage d'être deux fois plus rapide... et d'éviter des sur-segmentations sans conséquences pour la qualité des mouvements calculés, mais perturbant la lisibilité des résultats.

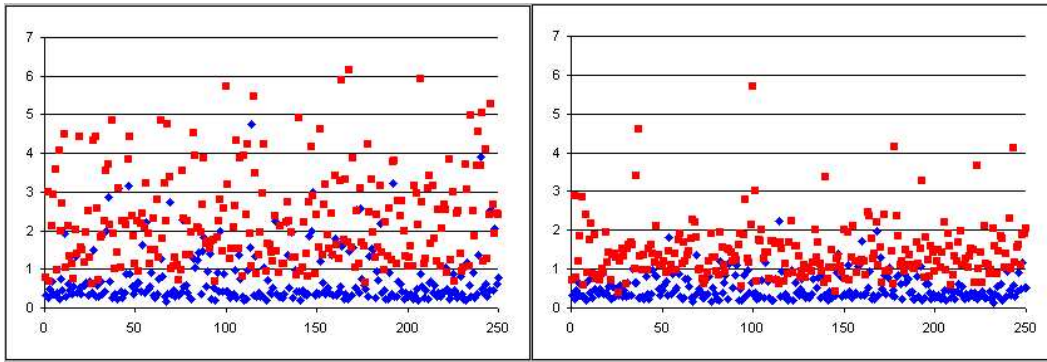


FIG. 5.5: Erreurs (mesurées avec l'EQM à gauche et l'EEM à droite) commises pour 250 générations d'images. Les points bleus correspondent aux images diagnostiques avec 20% de diffusé, et celles de droite à des examens diagnostiques avec 20% de diffusé.)

TAB. 5.3: Moyenne de l'erreur d'estimation pour différents niveaux de bruit et différentes variations temporelles des mouvements successifs (avec 20% de diffusé).

	Métrique	Variation temporelle des mouvements successifs			
		0%	10%	20%	30%
Graphie $\sigma = 10$	EQM	0,62	0,75	0,86	0,89
	EEM	0,39	0,46	0,57	0,70
	RF	6,61	6,96	7,47	7,88
Fluoroscopie $\sigma = 20$	EQM	1,90	2,00	2,10	2,20
	EEM	1,10	1,23	1,28	1,40
	RF	8,86	9,15	9,47	10,02

Relaxation de la contrainte de constance temporelle. Nous étudions au tableau 5.3 l'impact de violations de l'hypothèse de constance temporelle du mouvement sur la qualité des estimations. Nous reproduisons l'expérience décrite dans le chapitre précédent en 4.3.1.2 en générant pour chaque couche deux modèles de mouvement proches simulant une variation temporelle de mouvement.

Les résultats mettent en évidence comme dans les deux chapitres précédents une dégradation progressive de la qualité des modèles de mouvement estimés, et montrent donc que le respect exact de l'hypothèse n'est pas crucial pour la méthode proposée.

Comparaison avec l'estimation mono-couche. Enfin, nous confrontons la méthode d'estimation des mouvements transparents en situation de bi-transparence à deux estimateurs *mono-couches*, ne prenant pas en compte la transparence. En effet, les images cliniques que nous considérons contiennent de larges régions homogènes, sur lesquelles l'hypothèse de la conservation de l'intensité est approximativement valide.

Il est certain que les méthodes ne prenant pas en compte la transparence perdront de l'information puisqu'elles ne recherchent qu'un seul mouvement par point, mais peut-être seront elles en mesure de calculer le déplacement correspondant à la couche la plus texturée.

Nous mesurons leur performance en calculant le RF monocouche

$$RF = \frac{\sqrt{\sum_{\mathbf{p} \in \mathcal{S}-T} (I(\mathbf{p}) - I(\mathbf{p} + \mathbf{u}))^2}}{N} \tag{5.10}$$

et l'EQM entre le déplacement estimé et le déplacement de la vérité terrain *qui lui est le plus proche*, en chaque point.

Pour comparer des approches également adaptées aux images cliniques, nous avons construit une méthode d'estimation mono-couche qui a la même structure que celle que nous avons présentée dans les chapitres 3, 4 et 5. Elle se déroule en trois étapes :

- une mise en correspondance par blocs mono-couche (block-matching classique) ;
- une extraction des modèles affines simplifiés par une transformée de Hough. Elle est suivie d'un étiquetage des blocs par le modèle qui minimise le résidu mono-couche⁶ ;
- une estimation paramétrique globale, basée sur la minimisation d'un résidu mono-couche.

Nous pensons que cette méthode d'estimation mono-couche est compétitive puisqu'elle utilise les techniques robustes au bruit qui ont fait leurs preuves dans le cadre bi-couche. La dernière étape (l'estimation paramétrique) est facultative car nous ne sommes pas certains qu'elle améliore les modèles estimés.

Nous rassemblons les résultats de ces simulations sur des images graphiques et fluoroscopiques (avec 20% de diffusé) dans le tableau 5.4. Ils mettent clairement en évidence le fait que les méthodes mono-couches ne sont pas en mesure d'estimer de manière fiable les mouvements puisqu'elles ne modélisent pas la physique de formation de l'image.

Par contre, nous verrons au chapitre 7 qu'elles permettent de capter les déplacements des objets les mieux contrastés, ce qui peut être suffisant pour certaines applications.

TAB. 5.4: Précision de l'estimation jointe de mouvement et de segmentation en situation de bi-transparence distribuée, pour différentes méthodes d'estimations : celle développée dans ce chapitre, et son équivalent mono-couche, avec ou sans estimation paramétrique. Les images utilisées sont des images graphiques et fluoroscopiques, avec un diffusé de 20%.

Méthode d'estimation	Graphie		Fluoroscopie		Temps de calcul (s)
	EQM	RF	EQM	RF	
Bitransparence distribuée	0,62	6,61	1,90	8,86	110
Monocouche non paramétrique	2,57	11,2	3,03	11,5	2
Monocouche paramétrique	2,67	10,6	3,15	11,1	26

5.5.2 Résultats sur séquences réelles

Nous présentons maintenant des exemples d'estimation de mouvement en situation de bitransparence distribuée sur quatre séquences d'images réelles : une graphie et trois fluoroscopies. Ils illustrent la qualité de l'estimation des mouvements, mais mettent en évidence également la difficulté d'interprétation

⁶Nous n'utilisons bien sûr pas de terme de lissage car cette approche n'appréhende (au mieux) en chaque pixel qu'une des deux couches réelles.

des résultats, notamment quand la méthode conduit à une sur-segmentation de l'image. Le site <http://www.irisa.fr/vista/Equipe/People/Auvray/-DemoBiDistributedTransparency.html> donne accès aux séquences de trois images correspondant aux estimations présentées dans ce document, et présente également la séquence entière en question avec les modèles de mouvements qui lui sont associés.

Nous représentons les champs de mouvements des différentes couches en couleur, en multipliant la taille des vecteurs par 4 par rapport au déplacement estimé sur les images 864×864 , et par 2 sur les images 576×576 . Rappelons que dans tout ce travail, les vecteurs de déplacement calculés sont orientés vers le passé.

Séquence graphique 401. La séquence 401 présentée à la figure 5.6 se réfère à un examen diagnostique cardiaque au cours duquel une injection a lieu. Les images traitées sont de dimension 576×576 pixels. On y observe le cœur superposé à la colonne vertébrale, et particulièrement visible par sa frontière en haut à droite de l'image. Outre le fond statique, les mouvements observés sont les battements cardiaques (qui entraînent les poumons), et ceux des artères coronaires (surtout visibles après l'injection).

Deux couches seulement sont nécessaires pour expliquer les mouvements de l'image (si l'on excepte celui des artères coronaires injectées). À certains instants (comme le premier présenté), la méthode n'introduit que les deux couches nécessaires, mais il arrive qu'un nombre plus grand de couches soit introduit. Une sur-segmentation des couches a alors lieu (comme sur l'image de droite) mais les modèles de mouvement correspondants sont proches (du moins sur les régions texturées de l'image). Les estimations ainsi fournies sont tout à fait exploitables pour les mouvements en chaque point de l'image, même si la sur-segmentation rend l'analyse plus délicate.

Lorsque les artères injectées ont un mouvement incohérent, elles sont ignorées (la plupart du temps) comme sur l'image de droite. Quand leurs mouvements sont plus réguliers, il arrive qu'ils soient appréhendés. C'est le cas sur l'image centrale de la Fig.5.6 : la couche représentée en vert correspond aux artères coronaires dans sa partie supérieure, et aux battements cardiaques dans sa partie inférieure. La transition entre les deux est possible car la zone intermédiaire est peu texturée.

Il est délicat d'évaluer la qualité des estimés sur l'intégralité de la séquence. On peut tabler sur 85% de mouvements correctement estimés avant l'injection et 70% pendant, ce qui est très satisfaisant.

Séquence fluoroscopique 6. Nous présentons les résultats du traitement de la séquence fluoroscopique 6 à la figure 5.7 (cette séquence est composée de 50 images de dimension 576×576 pixels). Nous l'avons également traitée avec la méthode d'estimation de mouvement adaptée à la situation de bitransparence simple en 3.5.2. La séquence comprend en fait trois couches transparentes ; elle fait intervenir un fond statique, les mouvements respiratoires (des poumons et du diaphragme) et les battements cardiaques.

Quand deux mouvements seulement sont présents, ils sont estimés correctement comme sur la première image, au prix d'une éventuelle sur-segmentation. Lorsque les trois couches sont animées de mouvements différents, nous sommes en mesure de les estimer toutes les trois (et non pas seulement les deux qui dominent comme en 3.5.2). C'est le cas de la deuxième image, sur laquelle cette couche bleue correspond aux battements cardiaques. Seuls les blocs sur lesquels la couche (co-)domine lui sont attribués.

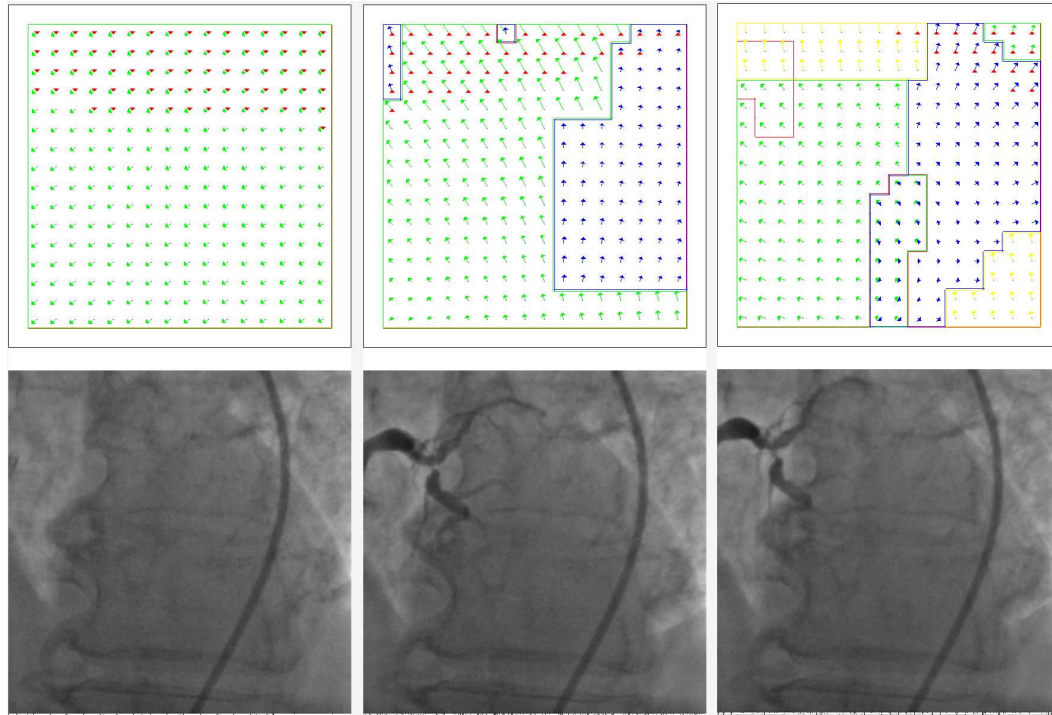


FIG. 5.6: Estimation du mouvement de la séquence diagnostique 401, en situation de bi-transparence distribuée (voir texte principal).

L'estimateur a su capter le mouvement de l'artère dans laquelle se situe le cathéter à la troisième image, et lui attribuer une couche dont les limites correspondent bien à l'outil interventionnel. Ce résultat très convaincant n'est malheureusement pas représentatif du comportement normal de la méthode : si elle arrive souvent à appréhender les mouvements des outils interventionnels ou des injections (quand ils se déplacent de manière souple), il est rare que la couche correspondante soit délimitée de manière aussi fiable.

Une autre supériorité de la méthode introduite dans ce chapitre par rapport à la bi-transparence simple du chapitre 3 est de pouvoir se limiter à une couche unique (si la transformée de Hough n'extrait qu'un modèle significatif). Cela contrôle le risque de voir apparaître des modèles de mouvements "fictifs" (voir la dernière image présentée). Cela dit, ce risque n'est pas complètement écarté puisque nous avons choisi de ne pas intégrer à la méthode le module de détection des configurations autre que bi-couches.

Sur l'ensemble de la séquence, l'estimation ne semble faillir que lorsque l'hypothèse de constance temporelle des mouvements est faussée, ce qui signifie que les estimés sont intéressants dans 90% des cas.

Séquence fluoroscopique 10. La séquence fluoroscopique 10 (de taille 864×864 pixels, aussi traitée en 4.3.2) illustre les problèmes liés à la sur-segmentation. Il semble que les réglages utilisés, pertinents sur des images 576×576 , ne sont pas aussi appropriés quand il s'agit de traiter des images 864×864 : l'introduction de nouveaux modèles de mouvement est acceptée trop facilement étant donné le nombre de blocs présents. Les premiers modèles

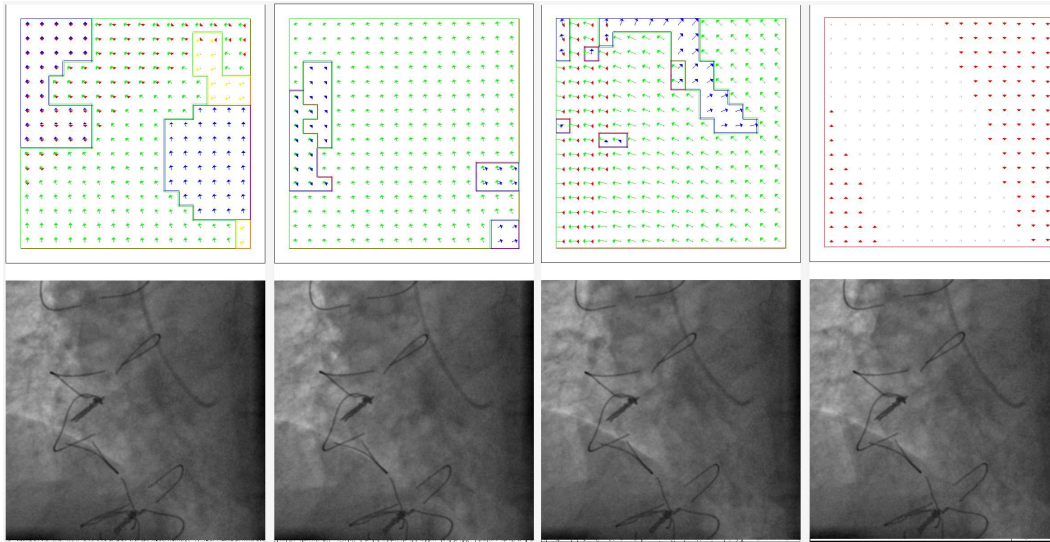


FIG. 5.7: Estimation du mouvement de la séquence fluoroscopique 6, en situation de bi-transparence distribuée (voir texte principal).

introduits sont souvent les plus pertinents (les couleurs des couches permettent de savoir dans quel ordre ils ont été extraits de la matrice de Hough : rouge, puis bleu, vert, jaune, violet).

Si un meilleur réglage aurait probablement mené à des résultats visuellement plus acceptables, il est important de remarquer que les modèles de mouvement estimés restent pertinents : ils correspondent bien aux déplacements ayant lieu sur les blocs texturés de l'image. La lisibilité des résultats est certes rendue délicate d'une part par certains champs "fictifs" et d'autre part par l'étiquetage de blocs non texturés, mais les mouvements du cœur, des poumons et du diaphragme sont la plupart du temps correctement appréhendés. Étant donné le manque de lisibilité des modèles de mouvement estimés, nous n'avons pas pu évaluer la qualité des estimations sur l'ensemble de la séquence. Seule leur utilisation dans le cadre de l'application visée (en ce qui nous concerne, le débruitage) permettrait de conclure avec précision sur leur pertinence.

Séquence fluoroscopique 19. Enfin, la séquence fluoroscopique 19 (images de taille 576×576 pixels) a été traitée. Elle a déjà été décrite au paragraphe 4.3.2. L'expérimentation conduit à répéter certaines observations :

- Quand deux couches seulement sont en mouvement, la méthode peut ou bien n'en introduire que deux (deuxième image), ou bien sur-segmenter certaines d'entre elles, auxquelles elle attribue des modèles de mouvement proches (quatrième image).
- La méthode d'estimation permet également d'expliquer des mouvements secondaires qui auraient été gommés par une approche considérant une situation de bi-transparence simple (première image).
- Elle produit par contre parfois des champs de déplacements très confus visuellement, même si les mouvements pertinents sont la plupart du temps correctement calculés

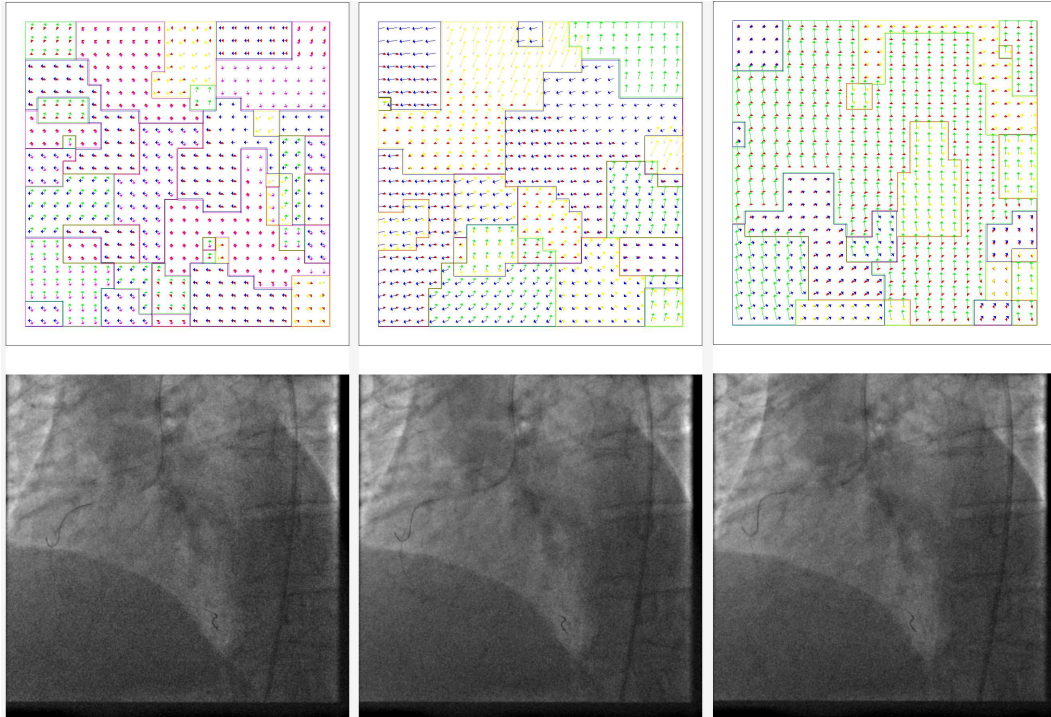


FIG. 5.8: Estimation du mouvement de la séquence fluoroscopique 10, en situation de bi-transparence distribuée (voir texte principal).

sur les régions texturées (troisième image).

Globalement, les mouvements semblent correctement estimés sur à peu près 80% de la séquence.

5.6 Conclusion

Nous avons traité dans ce chapitre le cas général des séquences d'images issues d'exams par rayons X, à travers le concept de la bitransparence distribuée. La méthode proposée est entièrement automatique : elle ne nécessite pas que l'utilisateur fournisse la segmentation de l'image en couches. Elle se base sur un schéma joint d'estimation des mouvements transparents et de segmentation spatiale par le mouvement des couches transparentes de l'image. Nous avons tenu à ne pas découpler ces deux aspects car ils sont fortement interdépendants. Nous avons pour cela introduit une formalisation markovienne du problème qui s'appuie sur la minimisation itérative d'une fonctionnelle faisant intervenir un terme (robuste) d'attache aux données et un terme de lissage. Nous avons expliqué comment régler la pondération entre ces deux termes indépendamment de l'image particulière traitée. La minimisation en elle-même s'effectue de manière itérative, par la méthode des moindres carrés pondérés itérés quand l'étiquetage est fixé pour estimer les modèles de mouvement, et par la technique ICM quand les modèles de mouvement sont fixés. Elle est initialisée avec les modèles de mouvement affines simplifiés qu'une transformée de Hough extrait des

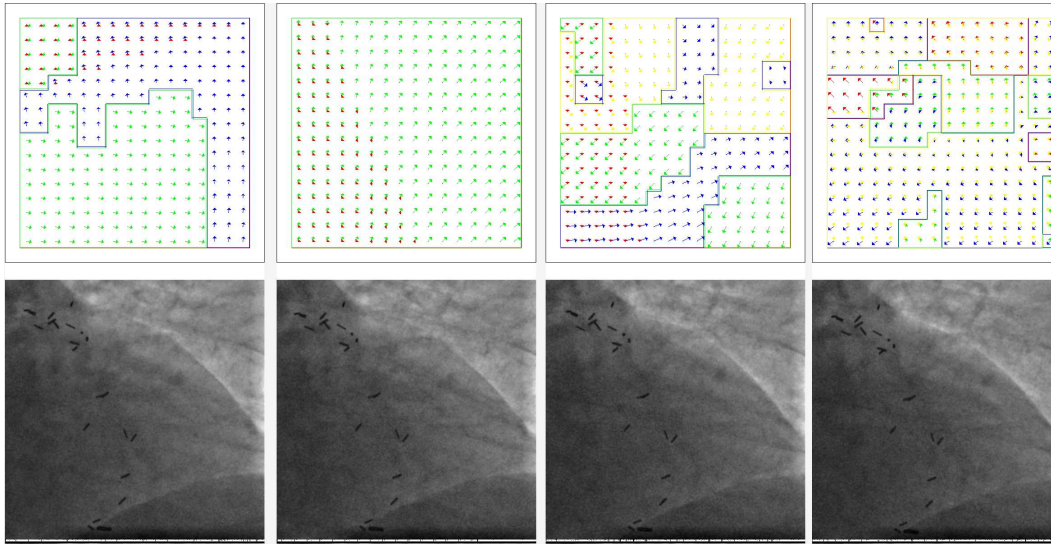


FIG. 5.9: Estimation du mouvement de la séquence fluoroscopique 19, en situation de bi-transparence distribuée (voir texte principal).

champs de déplacement calculés par mise en correspondance par blocs. Cette initialisation permet également de déterminer le nombre de couches présentes dans l'image. Nous pouvons ensuite réduire ce nombre, en fusionnant deux couches aux mouvements proches ou en faisant disparaître une couche trop peu représentée. En réglant la transformée de Hough pour favoriser les sur-segmentations, il ne s'avère pas nécessaire d'utiliser de mécanisme d'ajout de couches.

L'approche proposée permet également d'estimer les mouvements présents sur les régions mono-couches (au prix de l'introduction d'un modèle fictif), et est robuste à la présence de régions à plus de deux couches (si elles sont peu nombreuses). Il est donc inutile de chercher à détecter ces deux configurations, tâche très difficile sur le type d'images considéré.

La méthode se comporte de manière fiable sur les séquences synthétiques que nous avons utilisées. Elle se montre en particulier peu sensible à l'hypothèse de constance des mouvements sur deux intervalles de temps. La comparaison avec une approche mono-couche compétitive montre que la méthode d'estimation des mouvements transparents conduit à des précisions clairement supérieures.

En ce qui concerne les expérimentations sur séquences réelles, si l'on dépasse les problèmes de lisibilité des résultats qu'impliquent les éventuelles sur-segmentations, nous avons montré que la méthode jointe d'estimation et de segmentation de mouvements transparents permet de calculer les champs de mouvement des organes, et bien souvent des outils interventionnels, de manière satisfaisante.

Synthèse de la partie II

Dans cette deuxième partie, nous avons progressivement exposé le développement d'une méthode d'estimation des mouvements des organes adaptée à la problématique complexe considérée :

- La modalité d'imagerie par rayons X se traduit par des images **transparentes additivement**. Comme en particulier l'hypothèse de conservation de l'intensité des points le long de leurs trajectoires n'est plus vérifiée, les méthodes classiques d'estimation du mouvements ne sont pas applicables.
- Comme le rayonnement auquel sont exposés le patient et l'équipe médicale est aussi limité que possible, **les images obtenues sont difficiles** : elles sont bruitées, peu contrastées, et entachées d'un "voile", le diffusé.
- Le taux d'acquisition est élevé (30 images par seconde), et l'application visée à terme devra se dérouler en **temps réel**.

Approche suivie

Nous avons tout d'abord dressé un état de l'art des méthodes d'estimation du mouvement transparent. Les premières approches proposées historiquement tiraient profit de la robustesse d'estimateurs de mouvement ne prenant pas en compte la transparence, et ne sont donc pas adaptées à la situation de transparence totale que nous considérons. Les méthodes se plaçant dans l'espace de Fourier 3D sont pertinentes si les mouvements étudiés sont constants sur une dizaine d'images au moins, ce qui est incompatible avec le mouvement cardiaque.

Nous nous sommes donc appuyés sur la **généralisation de la conservation de l'intensité pour le cas de la transparence**, qui implique deux hypothèses :

- Les mouvements doivent être **localement translationnels**. Cette hypothèse n'est pas critique dans notre cas car les mouvements des organes sont faiblement divergents, et les images que nous considérons contiennent de larges plages homogènes.
- Les mouvements doivent être **constants sur deux intervalles de temps successifs** (ou sur n intervalles de temps successifs dans le cas de transparence impliquant simultanément n couches). Nous avons montré que les méthodes que nous avons définies sont robustes à des violations mesurées de cette hypothèse.

Par contre, sur 5 à 20% des instants considérés, le battement cardiaque se traduit par une rupture forte du mouvement. L'estimation est alors erronée à ces dates, et une perspective de notre travail sera de les détecter pour les traiter spécifiquement.

Développement d'une méthode robuste

Il nous fallait dès lors dériver de (2.39) une méthode particulièrement robuste au haut niveau de bruit des images (dont les caractéristiques sont résumées dans la Tab.5.5). Pour contraindre le plus possible notre problème, nous avons cherché à mettre à profit la régularité des mouvements anatomiques en faisant l'hypothèse majeure suivante :

- **les mouvements des organes sont supposés suivre un modèle affine.**

Cette hypothèse n'est bien évidemment pas une modélisation exacte des mouvements anatomiques, ne serait-ce que parce que nous observons la projection $2D$ de mouvements $3D$. Cela dit, elle semble raisonnable vis à vis des déplacements observés sur les examens réels, qu'il s'agisse de la respiration (translation du diaphragme, dilatation des poumons) ou du battement cardiaque (en première approximation).

En outre, l'approche la plus complète que nous proposons segmente l'image en couches, étant entendu qu'une couche se définit par son mouvement (et non par des considérations anatomiques). Autrement dit, elle est à même de sursegmenter certaines anatomies quand les mouvements qui les animent sont trop complexes pour être expliqués de manière satisfaisante par un modèle affine unique.

Nous avons développé autour de cette hypothèse une méthode d'estimation des mouvements transparents particulièrement robuste, que nous avons d'abord mise au point sur des séquences contenant deux couches exactement. Le schéma d'estimation est basé sur une **estimation paramétrique robuste dans un formalisme de Gauss-Newton**.

Nous avons étudié avec attention son initialisation. Nous avons finalement retenu une procédure en deux temps :

- calcul d'un champ de déplacement par **mise en correspondance par blocs sur un critère prenant en compte la transparence**.
- extraction des deux modèles affines simplifiés dominants par **transformée de Hough**.

Enfin, grâce à de nombreux tests sur des images synthétiques réalistes, nous avons choisi avec soin les différents paramètres de la méthode, à commencer par les prétraitements imposés aux images avant estimation des mouvements.

Extension au cas général

Nous avons étendu cette méthode au cas général d'images comptant un nombre indéfini de couches. Nous avons alors posé une hypothèse supplémentaire :

- si le nombre de couches présentes globalement dans l'image est libre, nous postulons cependant qu'il en existe au plus deux en chaque point. Nous avons nommé une telle configuration la **bitransparence distribuée**.

Cette hypothèse est née de l'observation des séquences issues d'examens cliniques, sur lesquelles il est rare que trois couches *significatives* soient simultanément présentes dans la même région. Il est cependant clair qu'il s'agit d'une hypothèse simplificatrice, et qu'il arrive que les trois mouvements majeurs (fond statique, respiration et battement cardiaque) soient présents dans une même partie de l'image. Toutefois, nous avons établi sur de nombreux exemples que la méthode d'estimation définie prend les bonnes décisions

dans ces situations critiques en se privilégiant les couches dominantes.

Nous avons dans un premier temps étendu notre approche au cas de séquences **en situation de bitransparence distribuée à segmentation donnée**. Les caractéristiques de la méthode sont résumées au Tab.5.5. L'estimation paramétrique robuste des mouvements découle du cas de la bi-transparence simple, mais l'étape d'initialisation est différente. Elle commence par une estimation de mouvement sur chaque *région bicouche*, d'où on extrait un modèle de mouvement pour chaque couche.

Cependant, cette approche suppose une interaction avec un utilisateur (qui doit définir la segmentation en couches) qui n'est pas envisageable pour l'application de débruitage que nous visons. Nous avons donc développé dans un deuxième temps une **méthode d'estimation et de segmentation jointes des mouvements transparents** (voir encore le tableau 5.5). Elle est basée sur une formulation markovienne qui se traduit par la minimisation itérative d'une fonctionnelle combinant entre attache aux données et contrainte de lissage.

Elle fait appel à une initialisation par mise en correspondance par blocs puis extraction des modèles de mouvement par transformée de Hough, qui permet également d'initialiser le nombre de couches présentes dans l'image. Cette détermination du nombre de couches, réglée pour sur-évaluer sa valeur réelle, est ensuite réduite lorsque certaines couches se révèlent être non significatives ou redondantes.

Adéquation au cas réel

Ces trois méthodes ont été testées de manière extensive sur des séquences synthétiques réalistes, ce qui a permis de régler de manière optimale leurs paramètres et de mieux comprendre leur fonctionnement. Surtout, elles ont été appliquées avec succès à une large base de séquences réelles. Des résultats convaincants ont été obtenus lorsque toutes les hypothèses d'utilisation des méthodes semblaient vérifiées, montrant leur robustesse au bruit et au diffusé réel, mais aussi leur **capacité à appréhender des mouvements réels** (qui ne sont ni exactement affines, ni parfaitement constants sur deux intervalles de temps successifs).

Elles ont aussi mené à des résultats intéressants dans des situations plus critiques :

- les régions en **transparence triple** sont appréhendées correctement ; les deux premières méthodes se concentrent sur les couches dominantes, et la dernière segmente l'image de façon à rendre compte des trois mouvements en présence, selon les régions où ils dominent.
- les régions en **transparence simple** voient leur (unique) mouvement correctement estimé. Par contre, un modèle fictif est introduit. Il est sans conséquence pour l'application de débruitage mais gêne la lisibilité des résultats.
- les **outils interventionnels** sont ignorés sur les régions bicouches (sauf quand ils sont très contrastés, comme les cathéters injectés), et leur mouvement est même souvent estimé sur les régions mono-couches. Comme ils se déplacent de manière plus abrupte que les organes, cette estimation n'est fiable que quand leur mouvement respecte approximativement l'hypothèse de conservation temporelle du mouvement.

Enfin, nous nous sommes interdits lors du développement de ces approches l'utilisation de techniques trop lourdes en temps de calcul, de sorte que leur implantation en temps réel n'est pas irréaliste à terme. Pour l'heure, et sans aucune optimisation des calculs, l'estimation en bitransparence simple prend 20s environ, en bitransparence distribuée à segmentation donnée 1 min, et l'approche la plus complète dure 2 minutes (pour une image 288×288 pixels).

TAB. 5.5: Récapitulatif des trois méthodes proposées d'estimation des mouvements transparents.

Configuration	Problème posé	Méthode	Résultats
Bitransparence simple	Deux mouvements affines globaux	<ul style="list-style-type: none"> □ Mise en correspondance transparente par blocs (multirésolution). □ Extraction par transformée de Hough des deux modèles affines simplifiés dominants. (Incrémentation continue de la matrice d'accumulation, recherche hiérarchique des modèles). □ Estimation paramétrique robuste (schéma de Gauss Newton impliquant une linéarisation du second ordre) 	<p><i>Images synthétiques</i> Graphie : EQM 0,27 pixels. Fluoroscopie : EQM 0,82 pixels.</p> <p><i>Images réelles</i> Très convaincants.</p>
Bitransparence distribuée à segmentation donnée	Mouvements affines des n couches.	<ul style="list-style-type: none"> □ Estimation en bitransparence simple sur chaque région bicouche (sans l'étape paramétrique) □ Attribution des modèles de mouvement des régions bitransparentes aux couches □ Estimation paramétrique robuste sur toute l'image. 	<p><i>Images synthétiques</i> Graphie : EQM 0,70 pixels. Fluoroscopie : EQM 1,81 pixels.</p> <p><i>Images réelles</i> Convaincants.</p>
Bitransparence distribuée sans segmentation prédéfinie	Nombre de couches n , segmentation de l'image en couches, modèles de mouvement (affines) des couches	<ul style="list-style-type: none"> □ Mise en correspondance par blocs avec critère de transparence □ Extraction des modèles affines dominants (initialisation (de leur nombre) □ Minimisation itérative d'une fonctionnelle : <ul style="list-style-type: none"> - à étiquettes fixées, <p>estimation paramétrique des mouvements.</p> <ul style="list-style-type: none"> - élimination éventuelle de couches surnuméraires. - à modèles de mouvement fixés, réétiquetage par ICM. 	<p><i>Images synthétiques</i> Graphie : EQM 0,60, EEM 0,4. Fluoroscopie : EQM 1,90, EEM 1,1.</p> <p><i>Images réelles</i> Très convaincants</p>

Troisième partie

Débruitage compensé en mouvement avec transparence

Introduction de la partie III

Alors que la deuxième partie de ce document portait sur l'estimation des mouvements anatomiques dans les images fluoroscopiques, cette troisième partie se concentre sur une application possible de la mesure de ces mouvements : le débruitage de séquences d'images. La plupart des filtres de réduction de bruit existants exploitent la redondance de l'information dans les séquences d'images, qu'elle soit spatiale ou temporelle. Ils sont donc plus performants quand ils s'appuient sur des images compensées en mouvement car la séquence d'images ainsi construite est alors *stationnaire* temporellement. D'autre part, les séquences d'images fluoroscopiques que nous considérons sont très sensibles aux artefacts éventuels introduits par différents filtres :

- les objets d'intérêt étant peu contrastés, toute atténuation de l'information est critique ;
- les séquences d'images sont très bruitées, et acquises à une haute fréquence ($30Hz$), de sorte que l'observateur sera très sensible aux artefacts de coloration du bruit. En particulier, un bruit très corrélé spatialement et peu corrélé temporellement laisse une impression désagréable (“grains en mouvement désordonné”), même si l'écart-type du bruit résiduel après filtrage est faible.

Aussi, les filtres purement temporels sont particulièrement attrayants pour ce type d'image : les artefacts qu'ils introduisent sont mieux maîtrisés. Leur limitation principale réside dans leur puissance de débruitage restreinte (par opposition notamment à des filtres spatio-temporels au support plus large). Une compensation de mouvement efficace peut toutefois rendre ces filtres temporels plus performants.

Il n'est cependant pas possible de compenser les mouvements d'une séquence d'images transparentes comme on compenserait ceux d'une séquence d'images “classiques”. Il n'est envisageable ni de suivre la trajectoire des pixels de l'image composite (du fait de la transparence), ni d'effectuer le filtrage au niveau des couches, puisqu'elles ne sont pas *séparées*. **Il faut donc développer des méthodes spécifiques de compensation des mouvements transparents.**

Cette troisième partie s'articule en quatre chapitres.

- Le **chapitre 6** propose un bref **état de l'art des méthodes de débruitage de séquences d'images**, et insiste sur l'apport de la compensation de mouvement pour la plupart de ces approches.
- Les **chapitres 7 et 8** portent sur le **débruitage temporel pur**. Le chapitre 7 présente dans un premier temps un filtre temporel adaptatif réglé pour les séquences d'images cliniques, puis propose une méthode pour *compenser les mouvements trans-*

parents.

Une étude théorique des propriétés de ce filtre met en évidence le niveau restreint de débruitage qu'il permet. Nous proposons alors dans le chapitre 8 une méthode alternative, dite **filtre hybride**, qui combine les qualités de conservation de contraste du filtre temporel adaptatif compensé en mouvement, avec une puissance de débruitage plus convaincante.

- Enfin, le **chapitre 9** propose une extension de ces méthodes de compensation des mouvements transparents au **débruitage spatio-temporel**. Même si ces filtres introduisent pour le moment des bruits à l'esthétique peu satisfaisante, nous montrons qu'ils peuvent tirer profit d'une compensation du mouvement.

Chapitre 6

État de l’art du débruitage de séquence d’images

L’application majeure de l’estimation de mouvements transparents considérée dans cette étude est le débruitage de séquences d’images fluoroscopiques. Du fait du rayonnement X qui doit être limité pour protéger patient et médecin, les images interventionnelles sont entachées d’un bruit quantique important. Il mène à des CNR¹ inférieurs à 1 pour certains objets d’intérêt ! Il est alors crucial de proposer des filtres de réduction de bruit performants pour rendre l’image plus lisible et plus agréable. Ces filtres doivent non seulement réduire significativement le niveau de bruit des images, mais aussi préserver toute l’information pertinente et ne pas introduire d’artefacts gênants pour l’observateur.

Ce chapitre présente un bref état de l’art des méthodes de réduction de bruit dans des *séquences* d’images. Nous n’évoquerons pas les méthodes de débruitage exclusivement spatiales [Buades 05, Kervrann 06] car nous souhaitons profiter de la redondance temporelle, qui est particulièrement importante aux hauts taux de rafraichissement utilisés (15 et 30Hz). En outre, un débruitage traitant indépendamment les images les unes après les autres est susceptible d’introduire des problèmes d’esthétique du bruit délicats à régler (puisque de légères imperfections invisibles image par image peuvent donner lieu à des artefacts visuellement gênants lorsque la séquence est jouée à 30 images par seconde).

Cet état de l’art succinct se propose de présenter les différentes méthodes de débruitage adaptées aux séquences d’images, et il insistera sur le fait que chacune d’entre elles profite grandement d’une éventuelle compensation du mouvement. Nous nous sommes appuyés sur les états de l’art proposés dans [Kleihorst 94], [Brailean 95b], [Dekeyser 00] et [Boulangier 06a].

6.1 Filtres pondérateurs

Principe. Pour débruiter l’image au pixel \mathbf{p} et à la date t , ce type de filtre forme la moyenne pondérée des intensités sur un voisinage spatio-temporel $V(\mathbf{p}, t)$ de (\mathbf{p}, t) :

$$\hat{I}(\mathbf{p}, t) = \sum_{(\mathbf{p}', t') \in V(\mathbf{p}, t)} w(\mathbf{p}, t, \mathbf{p}', t') I(\mathbf{p}', t'). \quad (6.1)$$

¹Contrast To Noise Ratio : le rapport du contraste de l’objet d’intérêt avec l’écart-type du bruit.

Sur des régions *stationnaires* de la séquence en effet (que ce soit spatialement ou temporellement), l'image comprend aux pixels (\mathbf{p}, t) et (\mathbf{p}', t') des intensités de points physiques qui "se ressemblent". Cette situation se produit spatialement sur des régions homogènes, et temporellement sur des zones sans mouvement apparent. On peut alors approcher la valeur $\tilde{I}(\mathbf{p}, t)$ idéale de l'image non bruitée en effectuant une moyenne sur les pixels de la région stationnaire.

Par contre, une pondération de pixels appartenant à des objets différents se traduit par un floutage de l'image et une perte d'information. De nombreuses méthodes ont été proposées pour choisir des poids $w(\mathbf{p}, t, \mathbf{p}', t')$ offrant d'excellents compromis entre puissance de débruitage et conservation des détails.

Soulignons que quelle que soit la stratégie finale retenue, ce type de filtre est plus performant quand l'image est compensée en mouvement car alors la stationnarité temporelle est vérifiée en tout point.

Filtre de Wiener. Dans le cas du filtre de Wiener, les poids du masque de convolution sont calculés de sorte à minimiser la différence quadratique moyenne entre la séquence originale et la séquence débruitée². Son utilisation a été étendue du domaine du traitement du signal (où son utilisation est classique) à celui du traitement de l'image [Özkan 91, Erdem 92, Kokaram 98, Boo 98].

Cette approche souffre du fait qu'elle nécessite la connaissance des propriétés de l'image originale, et qu'elle se base sur un modèle inexact de l'image puisqu'elle fait l'hypothèse que le signal est stationnaire. Le résultat final du débruitage peut donc être peu convaincant. Par contre, des versions locales [Dekeyser 00] et adaptatives bien plus intéressantes en ont été dérivées, dont le filtre LLMMSE³ qui est souvent cité comme référence lors de comparaisons de filtres de réduction de bruit [Dekeyser 00, Lee 06].

Filtres adaptatifs. Les filtres pondérateurs se basant sur des voisinages locaux sont beaucoup plus intéressants en pratique. Ils peuvent impliquer des voisinages purement spatiaux, purement temporels, ou spatio-temporels. Ils se différencient par les modes de calcul des poids proposés. Il s'agit de donner des poids forts aux pixels de la zone stationnaire, tout en limitant autant que possible l'importance de pixels appartenant à des régions différentes.

Certains filtres se basent pour ce faire sur une segmentation préalable de l'image sur la base de ses niveaux de gris [Dekker 85, Kalivas 90] ou de son mouvement [Irani 94], et donnent un poids important aux pixels de l'objet (et faible aux autres). Ce type d'approche est à réserver à des cas particuliers pour lesquels la difficile tâche de segmentation peut être réalisée de manière fiable (grâce en général à un *a-priori* important).

Les méthodes adaptatives pour qui le poids $w(\mathbf{p}, t, \mathbf{p}', t')$ est fonction des niveaux de gris $I(\mathbf{p}', t')$ et $I(\mathbf{p}, t)$ (ainsi qu'éventuellement des positions relatives des pixels (\mathbf{p}', t') et (\mathbf{p}, t)) sont applicables de manière beaucoup plus large. Les critères proposés traduisent le fait que plus les deux mesures $I(\mathbf{p}', t')$ et $I(\mathbf{p}, t)$ sont éloignées, plus il est probable qu'elles correspondent à des objets différents et donc plus le poids $w(\mathbf{p}, t, \mathbf{p}', t')$ doit être

²Sous l'hypothèse que le signal est un processus stochastique de corrélation et autocorrélation connue.

³Local Linear Minimum Mean Square Error

faible. Inversement, plus $I(\mathbf{p}', t')$ et $I(\mathbf{p}, t)$ sont proches, plus il est probable que la différence observée soit due au bruit et plus la pondération doit être importante. À partir de cette idée, différentes fonctions décroissantes de $|I(\mathbf{p}', t') - I(\mathbf{p}, t)|$ ont été proposées [Dubois 84, Özkan 93, Tomasi 98].

Ces méthodes ont l'avantage d'être rapides et de permettre des débruitages intéressants à niveau de bruit modéré. Elles deviennent plus problématiques quand le niveau de bruit est important car la frontière entre variation typique due à un changement d'objet et variation due au bruit devient plus floue.

Ce type de raisonnement a été étendu récemment à des filtres adaptatifs par "patches", dans le domaine de l'image simple [Buades 05, Azzabou 06] ou de séquences d'images [Boulanger 06b, Boulanger 06c]. La probabilité que (\mathbf{p}', t') et (\mathbf{p}, t) relèvent du même "objet" y est alors évaluée sur la base de la ressemblance entre les *voisinages* (ou *patches*) de (\mathbf{p}', t') et (\mathbf{p}, t) (et non pas sur le seul niveau de gris de l'image en ces deux points). Ces méthodes offrent des résultats intéressants même à niveau de bruit élevé.

6.2 Filtres séquentiels.

Filtres de Kalman. Le filtre de Kalman donne une estimée optimale (au sens des moindres carrés) d'une variable aléatoire dont on connaît

- la loi d'évolution linéaire (et entachée d'un bruit gaussien de statistique connue),
- la loi d'observation linéaire (et entachée d'un bruit gaussien de statistique connue).

L'utilisation de ce filtrage a permis de nombreuses avancées dans le domaine du traitement de l'image. Il doit en partie son succès à sa structure récursive, qui procède en deux temps pour aboutir à l'estimation. Une prédiction \bar{I} de l'image est d'abord effectuée (sur la base de l'estimée précédente), puis une mise à jour a lieu pour intégrer l'information issue de la nouvelle mesure. (La notation \bar{X} dénotera d'ailleurs tout au long de ce document la *prédiction* et non la moyenne!)

Ce schéma a été repris à des fins de débruitage de séquences d'images [Cano 83, Woods 93]. Cependant, les violations de l'hypothèse de stationnarité lorsque des mouvements ont lieu posent problème en créant des artefacts extrêmement gênants. Une version adaptative du filtre de Kalman (simplifié) donne de meilleurs résultats perceptuellement [Katsaggelos 89].

Filtre récursif. Un autre filtre récursif (simple mais empirique) a également été souvent utilisé. Il s'appuie aussi sur une prédiction \bar{I} qui n'est autre que l'image débruitée à la date précédente ($\bar{I}(\cdot, t) = \hat{I}(\cdot, t - 1)$), puis effectue la mise à jour suivante :

$$\hat{I}(\mathbf{p}, t) = (1 - \alpha(\mathbf{p}, t))\bar{I}(\mathbf{p}, t) + \alpha(\mathbf{p}, t)I(\mathbf{p}, t) \quad (6.2)$$

$$= (1 - \alpha(\mathbf{p}, t))\hat{I}(\mathbf{p}, t - 1) + \alpha(\mathbf{p}, t)I(\mathbf{p}, t). \quad (6.3)$$

Ce filtre a l'intérêt d'être très simple à mettre en œuvre, de ne nécessiter le stockage que d'une image ($\hat{I}(\cdot, t - 1)$) et de ne coûter qu'une opération par pixel (outre les calculs dus à la mise à jour du poids). Enfin, il permet d'adapter le poids $\alpha(\mathbf{p}, t)$ à la configuration présente au pixel en question. Les contributions portant sur le sujet proposent différentes politiques d'adaptation du poids [McMann 78, Dubois 84, Crawford 82], qui ont en commun d'être des fonctions croissantes de l'erreur de prédiction.

Outre ses qualités de rapidité et de facilité d'implantation, cette méthode présente l'avantage très précieux pour le traitement d'images fluoroscopiques de n'introduire que peu d'artefacts. Les trois artefacts majeurs sont :

- la *perte de contraste* : un objet en mouvement sera atténué si on ne lui attribue pas un gain exactement égal à 1. Or, nous expliquerons dans le chapitre suivant qu'il est nécessaire sur des images très bruitées de garder un gain inférieur à 1 pour atténuer les réalisations du bruit correspondant à la queue de sa distribution (sans quoi il reste un bruit de type "poivre et sel"). Par conséquent, tous les objets en mouvement sont partiellement atténués.
- le *lag*. La perte de contraste de l'objet en mouvement n'est pas la seule conséquence d'un gain différent de 1 sur une région en mouvement : cela implique aussi la réintroduction d'une partie de l'objet aux positions où il était présent à la date précédente. Cet effet de "fantôme", comme celui de la perte de contraste, sont cependant acceptables perceptuellement lorsque le filtre est bien réglé.
- la "*queue de bruit*". Quand un objet se déplace et que le coefficient α est correctement réglé, il ne sera que peu filtré, comme le fond qu'il découvre. La scène présente donc un niveau de bruit plus important sur la zone découverte, et ce pendant quelques instants (le temps que le filtrage récursif réduise suffisamment son niveau de bruit). Ce phénomène de queue de bruit, peu gênant comparativement aux artefacts introduits par certaines approches spatio-temporelles, peut être gommé très rapidement lorsque les poids sont mis à jour en utilisant l'historique du débruitage en chaque pixel [Aach 96b].

Finalement, la seule véritable limitation de ce filtre pour le débruitage d'images fluoroscopiques vient de son pouvoir de débruitage limité. Le fait que nous souhaitons conserver des objets interventionnels de contraste faible nous conduit à régler le poids α de manière conservatrice (sans quoi ils disparaîtraient lors de leurs déplacements), et ce faisant limite le gain de SNR. Une compensation fiable des mouvements nous autoriserait à filtrer de manière plus agressive la séquence d'images compensée en mouvement sans risquer de perdre (trop) d'information. C'est ce que nous proposons de faire en tenant compte de la transparence dans le chapitre 7.

6.3 Filtres d'ordre

D'autres auteurs se sont intéressés aux filtres d'ordre, essentiellement parce que ces derniers sont connus pour préserver les contours. Si la classe des filtres d'ordre contient l'ensemble des filtres se basant sur une liste ordonnée des niveaux de gris des pixels du voisinage, seul le filtre médian a véritablement été utilisé à des fins de débruitage [Huang 81, Naqvi 86] (à l'exception notable de [Kleihorst 94]).

Il a l'inconvénient d'être coûteux en temps de calcul si le voisinage est large, d'être moins efficace en terme de puissance de débruitage qu'un filtre moyenné (puisqu'il est plutôt destiné à éliminer le bruit impulsionnel), mais surtout de risquer d'introduire des artefacts sur certains contours. En effet, si le filtre médian $3D$ est efficace pour préserver la plupart des contours puisqu'il n'introduit pas de flou, il peut conduire à effacer certains contours, notamment ceux qui se déplacent.

En réponse à ce problème, des filtres médians à plusieurs échelles ont été développés

[Arse 91, Alp 91]. Ils s'appuient sur une première série de médians orientés (dans des directions d'intérêt), puis gardent le médian des résultats de l'ensemble de ces filtres. Lorsque les directions d'intérêt sont bien choisies, on lutte efficacement contre l'effacement de contours dans ces directions.

Les filtres d'ordre $3D$ gagnent aussi à considérer des images compensées en mouvement. Ce ne sont cependant pas ceux qui sont le mieux adaptés à notre problématique à cause de leur puissance de débruitage limitée.

6.4 Restauration bayésienne

Évoquons également la classe des approches bayésiennes. Pour reconstruire l'image originale \tilde{I} , on adopte un formalisme statistique. On cherche à maximiser la probabilité *a posteriori* $p(\tilde{I}|I)$ de \tilde{I} sachant que nous avons observé I . D'après la règle de Bayes, cette probabilité est proportionnelle au produit de $p(I|\tilde{I})$ (la probabilité d'observer I en connaissant l'image originale \tilde{I}) et de $p(\tilde{I})$ (la probabilité de l'image \tilde{I}).

$$p(\tilde{I}|I) \propto p(I|\tilde{I})p(\tilde{I}) \quad (6.4)$$

Si $p(I|\tilde{I})$ se déduit directement à partir du modèle du bruit, il est moins évident d'établir $p(\tilde{I})$ puisqu'il faut construire un modèle de l'image le plus général possible.

Ce problème a été traité sur la base de considérations locales via les champs de Markov [Hong 89], et par une minimisation fonctionnelle [Geman 92]. Dans un cas comme dans l'autre, le modèle de l'image est délicat à établir, et la résolution pratique du problème est extrêmement coûteuse.

6.5 Filtrage dans des espaces transformés

Les succès du débruitage d'images fixes dans des espaces transformés, que ce soit par seuillage des transformées en cosinus discrets [Aach 96a] ou surtout des transformées en ondelettes [Donoho 94, Mallat 98], ont inspiré des travaux portant sur le débruitage de séquences vidéo.

Cependant, l'adaptation au cas " $2D$ plus temps" est moins directe qu'il n'y paraît :

- d'une part une représentation précise des discontinuités dans un espace à trois dimensions est bien plus complexe qu'en deux dimensions [Selesnick 03],
- d'autre part les dimensions spatiales et temporelles ont des propriétés différentes, de sorte qu'une représentation très pertinente spatialement n'est pas forcément bien adaptée à la direction temporelle.

En réponse au premier point, certains auteurs utilisent des espaces de représentations plus riches (et redondants), afin de pouvoir représenter les orientations des objets *et* leurs mouvements de manière efficace [Selesnick 03]. Ces travaux sont prometteurs mais encore à un stade préliminaire.

L'essentiel des approches utilisant des espaces transformés pour débruiter les vidéos traite différemment les dimensions spatiales et la dimension temporelle. On peut ainsi utiliser successivement un filtrage par seuillage d'ondelettes et un filtre adaptatif temporel récursif [Pizurica 03], ou effectuer ces deux tâches parallèlement pour finalement retenir la moyenne des deux approches [Dugad 01]. Dans [van Roosmalen 96], la particularité de l'axe temporel

est éliminée en compensant le mouvement, puis un filtrage spatial basé sur des pyramides de Simoncelli et un filtrage temporel par seuillage d'ondelettes sont utilisés. Quoiqu'il en soit, ce dernier type d'approches permet de meilleurs débruitages si la séquence est stationnaire temporellement, et donc si son mouvement a été compensé.

6.6 Filtrage par EDP

Les méthodes de filtrage spatial s'appuyant sur les équations aux dérivées partielles (EDP) ont également été adaptées aux séquences d'images. Il s'agit d'approches visant à minimiser des fonctionnelles du type :

$$E = \int_{\mathbf{p} \in \mathfrak{S}} (\hat{I}(\mathbf{p}) - I(\mathbf{p}))^2 + \lambda \int_{\mathbf{p} \in \mathfrak{S}} \Phi(\|\nabla \hat{I}(\mathbf{p})\|). \quad (6.5)$$

La fonction de lissage Φ initialement proposée (une fonction quadratique de $\|\nabla \hat{I}(\mathbf{p})\|$) ayant l'inconvénient de lisser les contours, [Perona 90] l'a substituée par un terme anisotrope pénalisant les variations orthogonales au gradient spatial d'intensité (pour les gradients forts). La minimisation de l'énergie résultante peut se faire en résolvant les équations d'Euler-Lagrange [Weickert 99, Tschumperlé 02] :

$$\hat{I}_{\tau=0} = I \quad (6.6)$$

$$\frac{\partial \hat{I}}{\partial \tau} = \operatorname{div} \left(\frac{\Phi'(\|\nabla \hat{I}(\mathbf{p})\|)}{\|\nabla \hat{I}(\mathbf{p})\|} \nabla \hat{I}(\mathbf{p}) \right) \quad (6.7)$$

où τ représente le temps de diffusion. Ce processus de diffusion lisse progressivement l'image le long des gradients significatifs. Les images résultantes sont très notablement débruitées⁴ et les contours pertinents sont préservés si les paramètres sont bien choisis. Elles peuvent par contre présenter une esthétique quelque peu "artificielle" (de type "dessin animé").

[Lee 98] a directement étendu ces approches aux séquences d'images, vues comme des objets 3D, mais il n'évoque pas sur les artefacts éventuels qu'implique sa stratégie. [Bourdon 04] explique notamment qu'une telle approche peut mener à débruiter différemment des régions voisines dans des zones homogènes (ce qui perturbe l'observateur), et propose de découpler la diffusion spatiale de la diffusion temporelle pour minimiser cet effet. Sa méthode est en outre dotée d'un mécanisme de détection bas niveau des discontinuités temporelles, plus robuste au bruit que le simple gradient temporel. [Kornprobst 98] se prémunit contre les problèmes de discontinuités temporelles en ne filtrant que le fond statique de la séquence d'images. Il effectue conjointement la détection du fond et son filtrage dans un formalisme qui peut être vu comme une généralisation de la diffusion anisotrope.

Ces méthodes restent cependant incapables d'effectuer le processus de diffusion le long de structures en mouvement d'amplitude importante. Dans les travaux de [Uttenweiler 03], le mouvement des objets est appréhendé, et intégré dans la diffusion, par l'analyse de tenseurs de structure. Cependant, ces travaux se concentrent sur les images particulières de microscopie par fluorescence, très bruitées mais dont les objets d'intérêt linéaires se prêtent bien à une analyse par tenseur de structure. D'autre part, l'approche perfectionnée de [Mikula 04] estime conjointement les mouvements et l'intensité par diffusion anisotrope

⁴Certaines méthodes revendiquent une diminution de l'écart-type du bruit d'un facteur 4.

des surfaces de niveaux. Cette méthode est mise en œuvre par éléments finis.

Le filtrage par EDP est donc susceptible de donner des résultats intéressants, mais la dimension temporelle résulte en des “contours” notablement discontinus dans le temps, dont la spécificité doit être prise en compte. L'étude des artefacts introduits par le filtrage anisotrope utilisé sur une séquence d'images en est à ses débuts, et le temps de calcul impliqué par la résolution itérative de la diffusion est problématique pour des applications en temps réel. Enfin, ces méthodes présentent généralement un grand nombre de paramètres à régler.

6.7 Le problème de la mesure de la qualité du débruitage

Diverses mesures permettent de mesurer de manière quantitative le gain apporté par le filtre réducteur de bruit à partir d'une vérité-terrain. Il peut s'agir :

- de la réduction de l'écart-type du bruit ;
- de l'augmentation du SNR (rapport signal-à-bruit) ou du CNR (rapport contraste d'un objet d'intérêt à l'écart-type du bruit) ;
- du PSNR, exprimé en décibels :

$$PSNR = 20 \log \frac{255}{\sqrt{\frac{1}{|\mathfrak{S}|} \sum_{\mathbf{p} \in \mathfrak{S}} (\hat{I}(\mathbf{p}, t) - \tilde{I}(\mathbf{p}, t))^2}}. \quad (6.8)$$

Quoiqu'il en soit, ces différentes mesures quantifient l'écart moyen à l'image idéale. C'est un indicateur intéressant, mais il ne rend pas compte des aspects cruciaux que sont la perte de contraste, ni l'esthétique du bruit.

Ce dernier terme a trait à une vérité peu étudiée, mais très importante en pratique : la corrélation du bruit (temporelle ou spatiale) a une influence majeure sur la perception qu'en a un observateur. Ainsi, un bruit fortement corrélé spatialement mais indépendant temporellement donne l'impression agrégé en “grains” se déplaçant chaotiquement. Il pourra être beaucoup plus perturbant qu'un bruit blanc d'écart-type triple !

C'est pourquoi l'observation des images débruitées est une étape incontournable pour juger de la qualité d'un filtre, surtout quand il s'agit de séquences d'images, particulièrement sensibles aux problèmes d'esthétique du bruit. Par contre, cette observation a l'inconvénient d'être éminamment subjective (et le support papier s'y prête mal!).

Un moyen intéressant de visualiser les artefacts éventuels et le *bruit de méthode* [Buades 05]. Il s'agit de former

$$\hat{I}(\mathbf{p}, t) - I(\mathbf{p}, t). \quad (6.9)$$

Si elle est vierge de tout lissage d'information utile, cette carte de différence entre l'image d'entrée et sa version débruitée ne doit laisser reconnaître aucun contour. Nous pouvons même nous permettre dans le cas de simulations à vérité-terrain connue de former la carte de *bruit résiduel*

$$\hat{I}(\mathbf{p}, t) - I_{ideal}(\mathbf{p}, t). \quad (6.10)$$

Elle permet non seulement de visualiser les artefacts mis en évidence par la carte de bruit de méthode, mais aussi de percevoir la corrélation spatiale du bruit de l'image traitée.

6.8 La question du mouvement

Les différentes approches décrites dans ce chapitre ont en commun d'être performantes dans les régions spatio-temporelles stationnaires de la séquence d'images. On y trouve une information hautement redondante, qui permet d'éliminer efficacement le bruit. La seule exception vient de la restauration bayésienne qui permet d'introduire d'autres *a-priori*, mais nous avons souligné combien cette approche était délicate à mettre en œuvre.

Si les nombreux travaux portant sur le débruitage d'images (fixes) ont permis le développement de filtres spatiaux exploitant la stationnarité des régions homogènes tout en conservant les détails informatifs, l'ajout de la dimension temporelle pose problème. Cela est paradoxal car la redondance de l'information y est particulièrement importante.

Ce phénomène s'explique d'une part par le fait que les artefacts potentiels introduits par un traitement temporel sont plus gênants que ceux qui viennent de traitements spatiaux, et d'autre part, parce que la détection de non-stationnarités temporelles implique l'analyse des mouvements de l'image [Bertero 88].

On comprend dès lors l'importance cruciale de la compensation du mouvement pour des applications de débruitage de séquences d'images [Kleihorst 94]. De fait, pratiquement toutes les méthodes récentes citées utilisent une compensation du mouvement pour améliorer la stationnarité temporelle. Les seules exceptions éventuelles sont les filtres adaptatifs par patch dans le cas où le voisinage de recherche est large⁵.

À ce titre, le cas particulier des méthodes *jointes* de restauration des images et de l'estimation de leur mouvement est intéressant. Elles se basent sur le constat que ces deux problématiques sont liées puisque l'estimation de mouvement est plus fiable sur des images débruitées, et le débruitage meilleur quand l'estimation du mouvement est plus précise. Toutefois, ces méthodes s'appuient sur un formalisme markovien généralement complexe [Brailean 95a], ou font des hypothèses simplificatrices fortes [Fan 99].

Si de très nombreuses contributions ont porté sur l'estimation et la compensation du mouvement dans des images vidéos classiques, elles ne peuvent pas s'appliquer aux images acquises par rayons X à cause du problème de la transparence. Même si ce problème de l'estimation du mouvement transparent est assez particulier, il a donné lieu à un certain nombre de contributions que nous avons énumérées dans le chapitre 2, ainsi qu'au développement des méthodes d'estimation dédiées aux images fluoroscopiques en situation de bi-transparence et bi-transparence distribuée que nous avons présentées dans les chapitres 3, 4 et 5.

Par contre, le problème de la *compensation* de mouvements transparents n'avait encore été traité avant notre article [Auvray 05c]. À la différence de la compensation de mouvement sur des images vidéo classiques, ce problème n'est en effet pas immédiatement résolu lorsque les mouvements sont estimés, comme nous allons le voir dans le chapitre 7.

⁵Ce qui a un lourd coût calculatoire.

Chapitre 7

Débruitage temporel compensé en mouvement de séquences d'images transparentes

Nous avons vu dans le chapitre précédent que le filtre adaptatif temporel récursif était très utilisé pour le débruitage de séquences d'images. Il est compact, rapide, et introduit peu d'artefacts pourvu que son gain soit réglé de manière adéquate. Nous introduisons dans une première section de ce chapitre un filtre adaptatif temporel adapté au traitement des images que nous considérons.

Lorsqu'il est réglé de manière conservatrice (pour conserver les objets d'intérêt de faible contraste), il a un pouvoir de débruitage limité. C'est pour dépasser cette limitation que nous présentons dans la deuxième section de ce chapitre une méthode originale de *compensation des mouvements transparents*.

L'étude des propriétés du filtre ainsi construit met en évidence sa capacité à conserver tous les détails de l'image, mais révèle aussi une limite de débruitage notable.

7.1 Filtre temporel récursif non compensé en mouvement

7.1.1 Principe

Rappelons la formule (6.2) sur laquelle se base ce filtre :

$$\hat{I}(\mathbf{p}, t) = (1 - \alpha(\mathbf{p}, t))\bar{I}(\mathbf{p}, t) + \alpha(\mathbf{p}, t)I(\mathbf{p}, t) \quad (7.1)$$

$$= \bar{I}(\mathbf{p}, t) + \alpha(\mathbf{p}, t)(I(\mathbf{p}, t) - \bar{I}(\mathbf{p}, t)) \quad (7.2)$$

avec $I(\mathbf{p}, t)$ l'image mesurée et $\bar{I}(\mathbf{p}, t)$ une prédiction de l'image à la date t (effectuée sans la connaissance de $I(\mathbf{p}, t)$). Rappelons également que $\hat{I}(\mathbf{p}, t)$ désigne l'image débruitée.

Lorsque la prédiction est fiable, ce filtre permet de conserver l'intégralité des détails de l'image, tout en atteignant un excellent débruitage (de par sa structure récursive).

- Dans un cadre classique, l'image prédite n'est autre que l'image débruitée à la date précédente : $\bar{I}(\mathbf{p}, t) = \hat{I}(\mathbf{p}, t - 1)$.

- Dans une version compensée en mouvement sans transparence, l'image prédite est l'image précédente déplacée par les mouvements estimés : $\bar{I}(\mathbf{p}, t) = \hat{I}(\mathbf{p} + \mathbf{w}(\mathbf{p}), t - 1)$.
- Nous détaillons le cas de la version compensée en mouvements transparents dans la partie 7.2.

7.1.2 Comportement sur régions stationnaires

Réglage du gain. Nous étudions dans un premier temps le réglage du gain *dans le cas idéal où la scène ne présente aucun mouvement*. Les variations de niveaux de gris entre des dates successives y sont dues au bruit seulement. Sous l'hypothèse d'indépendance du bruit entre deux instants, l'expression (7.1) induit la relation suivante sur les variances des bruits :

$$\sigma_{\hat{I}}^2(t) = (1 - \alpha(t))^2 \sigma_{\hat{I}}^2(t - 1) + \alpha(t)^2 \sigma^2. \quad (7.3)$$

où σ note l'écart-type du bruit des images acquises¹ et $\sigma_{\hat{I}}(t)$ l'écart-type du bruit résiduel sur l'image débruitée à la date t . Nous souhaitons choisir le gain $\alpha(t)$ qui minimise le bruit résiduel de l'image traitée² $\sigma_{\hat{I}}^2(t)$. Pour cela nous considérons la relation suivante :

$$\frac{d\sigma_{\hat{I}}^2(t)}{d\alpha(t)} = 2(\alpha(t)\sigma^2 + (\alpha(t) - 1)\sigma_{\hat{I}}^2(t - 1)) = 0. \quad (7.4)$$

Ce qui conduit à choisir la contrainte :

$$\alpha_{id}(t) = \frac{\sigma_{\hat{I}}^2(t - 1)}{\sigma^2 + \sigma_{\hat{I}}^2(t - 1)}, \quad (7.5)$$

Nous sommes donc en mesure de calculer itérativement les gains à appliquer à chaque instant, qui eux-même permettent d'évaluer par (7.3) le niveau de bruit attendu sur l'image traitée.

Nous montrons dans l'annexe F que le gain et la variance valent :

$$\alpha_{id}(t) = \frac{1}{t + 1} \quad (7.6)$$

$$\sigma_{\hat{I}}^2(t) = \frac{1}{t + 1} \sigma^2. \quad (7.7)$$

Remarquons au passage que nous atteignons bien le niveau de débruitage maximal théorique permis par l'utilisation de $t + 1$ mesures bruitées.

Application pratique. Nous donnons dans le tableau 7.1 les valeurs théoriques du gain et du niveau de bruit atteintes sur une séquence de 9 images statiques, ainsi que le débruitage réellement obtenu sur une séquence test synthétique statique (générée comme en 1.3.2, avec des mouvements nuls, un diffusé de 20% et un bruit de mesure $\sigma = 20$).

On note une **décroissance rapide du bruit** : il est divisé par 2 après 4 images et par 3 après 9 images (soit 1/3 de seconde à 30Hz). Cela dit, le gain décroît également

¹Qui est connu en pratique grâce à la mesure du rayonnement X reçu par le système d'imagerie.

²En d'autres termes, nous souhaitons choisir $\alpha(t)$ tel que \hat{I} soit un estimateur de l'image idéale de variance minimale.

TAB. 7.1: Evolution du débruitage sans adaptation

Image t	1	2	3	4	5	6	7	8
α_{id}	0.5	0.33	0.25	0.2	0.17	0.143	0.125	0.11
$\sigma_{\hat{I}}(t)/\sigma$ théorique	0.707	0.58	0.5	0.45	0.41	0.38	0.35	0.33
$\sigma_{\hat{I}}(t)/\sigma$ dans les simulations	0.72	0.58	0.51	0.46	0.43	0.40	0.38	0.36

très rapidement, ce qui implique que seule une petite partie de l'image mesurée est utilisée dans la pondération (7.1).

On comprend alors que ce type de réglage du filtre ne peut pas fonctionner sur des images réelles car très vite, les nouvelles informations mesurées ne seront quasiment pas (ou très peu) exploitées dans la reconstruction de l'image débruitée. **Les objets en mouvement risquent donc d'être effacés.**

7.1.3 Adaptation du gain

7.1.3.1 Filtrage adaptatif

Dans la pratique, il est donc nécessaire d'adapter le gain depuis cette valeur idéale, en **évaluant la probabilité que $\bar{I}(\mathbf{p}, t)$ soit bien une prédiction réaliste de $I(\mathbf{p}, t)$** . Cette adaptation, absolument cruciale quand le filtre est utilisé sans compensation de mouvement, reste nécessaire lorsque les mouvements sont compensés car la compensation n'est jamais parfaite. On peut par contre choisir une adaptation conduisant à un filtrage plus marqué dans ce dernier cas, car les non-stationarités sont beaucoup plus rares.

Il s'agit d'arbitrer entre deux hypothèses :

- $\bar{I}(\mathbf{p}, t)$ fournit la même mesure d'intensité que $I(\mathbf{p}, t)$, au bruit près,
- $\bar{I}(\mathbf{p}, t)$ est une mesure "physiquement" différente de $I(\mathbf{p}, t)$.

Si la probabilité de la première hypothèse nous est donnée par les statistiques du bruit, nous ne pouvons pas évaluer la deuxième sans utiliser un modèle de l'image. Par conséquent, il n'est pas possible d'arbitrer de manière statistiquement optimale entre ces deux hypothèses dans le cas général.

Nous sommes contraints de nous appuyer sur des heuristiques pour formuler une technique d'adaptation. L'état de l'art contient de nombreuses contributions sur le sujet [McMann 78, Dubois 84, Crawford 82, Aach 96b], qui ont en commun de proposer un gain

- croissant en fonction de l'erreur de prédiction,
- et décroissant en fonction du bruit de mesure.

7.1.3.2 Réglage pratique du gain

Dans le cas particulier des images fluoroscopiques sur 12 bits, nous souhaitons préserver les projections des outils interventionnels dans les images, dont les niveaux de gris sont typiquement dans la plage [20, 100].

Nous ne pouvons pour autant pas cesser de filtrer dès qu'une erreur de prédiction de plus de 20 niveaux de gris³ est obtenue (c'est-à-dire fixer $\alpha(\mathbf{p}, t) = \alpha_{id}(t)$ quand

³Ou plus précisément leur différence par rapport au fond.

$|I(\mathbf{p}, t) - \bar{I}(\mathbf{p}, t)| > 20$, et $\alpha(\mathbf{p}, t) = 1$ sinon). Le bruit a en effet un écart-type typique de l'ordre de 20, et cette stratégie de filtrage le réduirait trop faiblement. Nous préconisons une adaptation progressive du type de la courbe tracée en bleu sur la Fig.7.1.

Cependant, ce choix de gain entraîne un artefact de type “bruit poivre et sel”. Les réalisations du bruit correspondant à la queue de la gaussienne se traduisent par des erreurs de prédiction importantes, qui ne sont donc pas (ou peu) modifiées par ce type de gain. Ces cas de figure sont rares mais très perceptibles, et il est donc très gênant de les laisser trop longtemps inchangés à l'image.

C'est pourquoi nous privilégions un gain qui progresse plus lentement vers l'unité (courbe en rouge sur la figure 7.1), afin de réduire tous les niveaux du bruit. Cela implique bien sûr une perte de contraste global, mais c'est ce compromis qui est visuellement le plus acceptable.

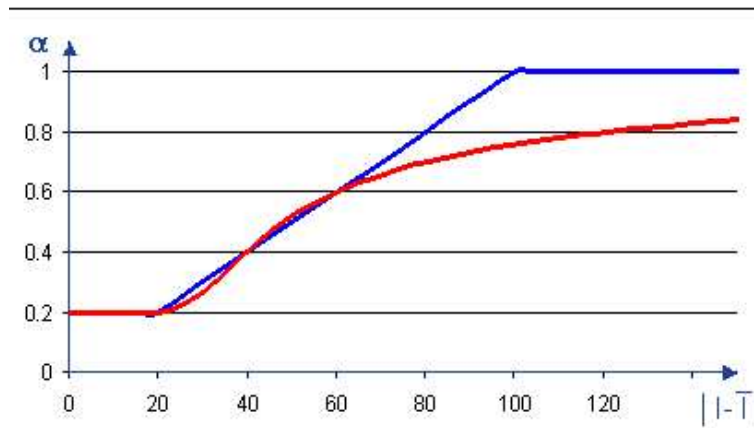


FIG. 7.1: Adaptation du gain du filtre temporel récursif en fonction de l'erreur de prédiction. Courbe bleue : approche progressive, courbe rouge : fonction prenant en compte l'effet “poivre et sel”.

Plus précisément, le gain choisi vaut

$$\alpha(\mathbf{p}, t) = \begin{cases} \alpha(\mathbf{p}, t) = \alpha_{id}(t) & \text{si } |I(\mathbf{p}, t) - \bar{I}(\mathbf{p}, t)| < 20 \\ \alpha(\mathbf{p}, t) = \frac{A|I(\mathbf{p}, t) - \bar{I}(\mathbf{p}, t)|^2 + B|I(\mathbf{p}, t) - \bar{I}(\mathbf{p}, t)| + C}{|I(\mathbf{p}, t) - \bar{I}(\mathbf{p}, t)|} & \text{si } 20 < |I(\mathbf{p}, t) - \bar{I}(\mathbf{p}, t)| < 40 \\ \alpha(\mathbf{p}, t) = 1 - \frac{D}{|I(\mathbf{p}, t) - \bar{I}(\mathbf{p}, t)|} & \text{si } |I(\mathbf{p}, t) - \bar{I}(\mathbf{p}, t)| > 40 \end{cases} \quad (7.8)$$

Cette fonctionnelle et ses coefficients ont été choisis de manière heuristique. Nous avons imposé sa forme dans les intervalles $[0, 20]$ et $[20, \infty[$, et choisis les coefficients A , B , C et D pour que la fonctionnelle soit \mathcal{C}^1 . Leurs valeurs sont rapportées dans le tableau 7.2.

En toute rigueur, puisque la valeur du gain varie spatialement, le niveau de bruit résiduel devrait aussi être fonction de \mathbf{p} : $\sigma_{\hat{I}}(t) = \sigma_{\hat{I}}(\mathbf{p}, t)$. À son tour, α_{id} devrait varier spatialement puisque son calcul (par la relation (7.5)) fait intervenir $\sigma_{\hat{I}}$: $\alpha_{id}(t) = \alpha_{id}(\mathbf{p}, t)$.

Seulement, nous ne pouvons pas évaluer de manière suffisamment précise ces valeurs en pratique. Nos expériences sur données synthétiques montrent en particulier que les estimés

TAB. 7.2: Valeur des coefficients nécessaires au calcul du gain

A	B	C	D
$\frac{1-\alpha_{id}}{40}$	$2\alpha_{id} - 1$	$10(1 - \alpha_{id})$	$30(1 - \alpha_{id})$

de $\sigma_{\hat{I}}(\mathbf{p}, t)$ calculés itérativement par la relation (7.3) (sur la base des poids choisis et de la connaissance du bruit de mesure) divergent rapidement des bruits résiduels réellement calculés.

C'est pourquoi nous avons décidé de négliger les dépendances spatiales de $\sigma_{\hat{I}}(t)$ (et de $\alpha_{id}(t)$). D'autre part, nous avons approché la valeur de l'écart-type du bruit résiduel au cours de la séquence par la formule suivante :

$$\sigma_{\hat{I}}(t) = \sigma(0,7 + 0,2e^{-\frac{t}{4}}) \text{ pour } t > 0 \tag{7.9}$$

établie sur la base de mesures effectuées sur une séquence simulée que nous avons traitée (Fig.7.2).

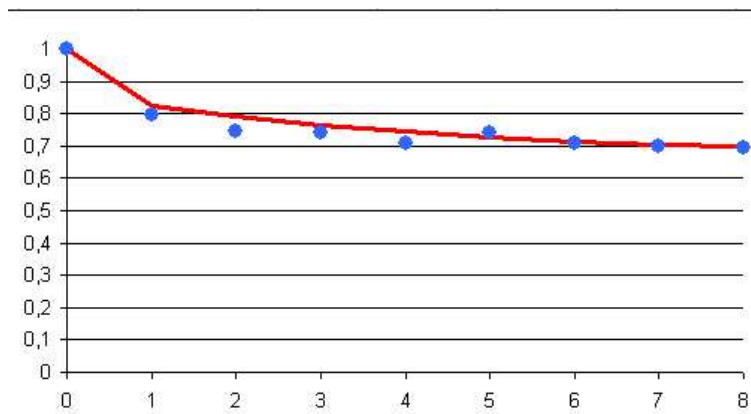


FIG. 7.2: Mesure du bruit résiduel sur une séquence d'images synthétiques traitée par le filtre adaptatif (tracé en bleu), modèle de bruit utilisé (tracé en rouge). Les ordonnées correspondent à l'écart-type du bruit normalisé ($\sigma_{\hat{I}}(t)/\sigma$), et les abscisses à l'indice de l'image débruitée.

7.1.3.3 Application pratique.

Nous allons comparer les performances du filtre temporel non adaptatif (FTNA) et du filtre temporel adaptatif (FTA) sur des séquences d'images synthétiques typiques d'un examen fluoroscopique, dans deux cas :

- celui pour laquelle la séquence est statique,
- celui pour laquelle la séquence et les couches se déplacent.

TAB. 7.3: Evolution du niveau de bruit d'une séquence statique synthétique traitée par les filtres adaptatifs (FTA) et non adaptatifs (FTNA)

n	1	2	3	4	5	6	7	8
$\sigma_{\hat{f}}(t)/\sigma$ FTNA	0.72	0.58	0.51	0.46	0.43	0.40	0.38	0.36
$\sigma_{\hat{f}}(t)/\sigma$ FTA	0.77	0.71	0.69	0.66	0.63	0.60	0.59	0.58

Sur images statiques Nous présentons les niveaux de bruit atteints pour une séquence synthétique *statique* traitée par le FTNA et le FTA dans le tableau 7.3. Nous y notons un débruitage notablement plus faible du FTA, surtout asymptotiquement, mais qui reste très intéressante. Il est d'ailleurs tout à fait logique que le filtre non adaptatif se comporte mieux puisque nous avons montré qu'il fournit les résultats optimaux pour le cas statique. Nous atteignons avec le filtre temporel adaptatif des débruitages de l'ordre de 45% en écart-type sur les zones statiques. Un gain réglé de manière à filtrer de manière plus agressive réduit plus fortement le bruit, mais atténue trop les contrastes d'objets d'intérêt en mouvement.

Sur images dynamiques. Nous avons construit une séquence de 9 images où les deux couches sont en mouvement constant : la couche abdominale en translation ($a_1 = 3.4$, $a_4 = 1.2$), et la couche cardiaque en mouvement homothétique ($a_1 = -2.3$, $a_2 = a_6 = 0.05$)⁴. Nous présentons la cinquième image de la séquence, traitée par les deux filtres, à la Fig.7.3. Les images présentées sont codées sur 12 bits, et ne sont pas traitées par les algorithmes d'affichage élaborés : nous ramenons simplement leur dynamique à $[0, 255]$. Nous avons fait ce choix pour que les algorithmes de visualisation ne viennent pas perturber la perception des résultats, mais cela se traduit par des images plus difficiles pour l'observateur. Le filtre récursif non adaptatif FTNA fournit une image mieux débruitée dans le sens où elle est plus lisse, mais complètement inacceptable puisque les détails sont effacés : les tissus du poumon, le cathéter et les agrafes notamment. Le filtre adaptatif FTA donne de bien meilleurs résultats : le niveau de bruit est notablement diminué comparativement à l'image acquise (de 30% en écart-type, soit un *PSNR* de 25.2), et les détails sont encore visibles même si certains sont atténués (tel que le cathéter).

Il faut noter que l'analyse qualitative de ces images *fixes* est perceptuellement à la défaveur des filtres présentés. Un observateur considèrerait que les séquences débruitées sont de bien meilleure qualité en observant les vidéos correspondantes puisque le bruit est beaucoup plus perceptible (et donc le gain en débruitage y est beaucoup plus appréciable!). De plus, l'oeil permet de suivre sans trop d'effort des objets qui apparaissent pourtant fortement effacés sur des images fixes.

C'est le FTA qui présentait le meilleur compromis débruitage/visibilité du point de vue du médecin jusqu'à une période récente. On est pourtant frappé de l'éloignement du résultat à l'image synthétique construite sans bruit, qui constitue la référence.

Nous proposons dans ce chapitre de compenser les mouvements des images, ce qui permettra d'utiliser plus efficacement le filtre temporel adaptatif et éventuellement d'opter pour

⁴Du fait de la composante a_1 non nulle, l'homothétie n'est pas centrée à l'origine.

des réglages de gain menant à un débruitage plus important.

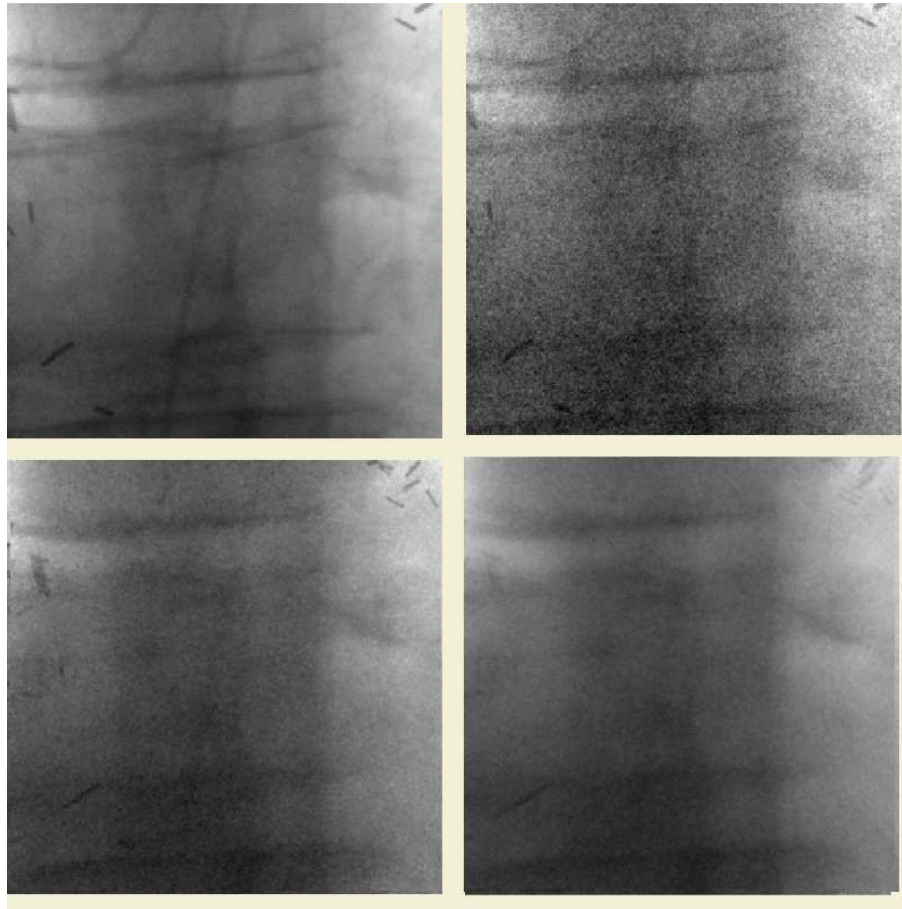


FIG. 7.3: Résultats du débruitage non compensé en mouvements transparents d'une séquence synthétique dont les deux couches sont en mouvement (la couche abdominale en translation et la couche cardiaque en mouvement homothétique). Les images présentées sont encodées en 12 bits et ne sont pas traitées par les algorithmes de présentation d'image. En haut : à gauche, l'image idéale non bruitée, à droite l'image mesurée. En bas : à gauche l'image débruitée par le filtre temporel récursif adaptatif (FTA), à droite l'image débruitée par le filtre temporel récursif non adaptatif (FTNA).

7.2 Compensation des mouvements transparents

7.2.1 Principe

La compensation du mouvement sur des images vidéo classiques (à savoir, sans transparence) consiste à calculer le champ de mouvement $\mathbf{w}(\mathbf{p}, t-1)$, puis à effectuer la prédiction :

$$\bar{I}(\mathbf{p}, t) = I(\mathbf{p} + \mathbf{w}(\mathbf{p}, t-1), t-1) \quad (7.10)$$

Ceci revient à suivre les trajectoires des pixels, et permet de se ramener à une situation stationnaire sur les régions non occultées si les mouvements sont bien estimés.

Il n'est pas possible de procéder de la sorte sur les séquences d'images fluoroscopiques puisqu'elles sont transparentes. En nous appuyant sur les estimés des mouvements transparents, nous pouvons certes suivre la trajectoire des pixels mais *nous ne connaissons pas leurs niveaux de gris* puisqu'il y a addition des différentes couches. Comme évoqué dans la partie 2.4 et l'annexe A, le problème de la séparation des couches est extrêmement délicat. Nous proposons une alternative en reformulant l'équation de contrainte du mouvement transparent (2.39) comme suit :

$$I(\mathbf{p}, t+1) = I(\mathbf{p} + \mathbf{w}_1, t) + I(\mathbf{p} + \mathbf{w}_2, t) - I(\mathbf{p} + \mathbf{w}_1 + \mathbf{w}_2, t-1). \quad (7.11)$$

En d'autres termes, lorsque les mouvements des deux couches sont connus, il est possible de *prédire* l'image à la date $t+1$ à partir des deux images précédentes. En reprenant les notations de ce chapitre, nous proposons l'équation suivante :

$$\bar{I}(\mathbf{p}, t) = I(\mathbf{p} + \mathbf{w}_1, t-1) + I(\mathbf{p} + \mathbf{w}_2, t-1) - I(\mathbf{p} + \mathbf{w}_1 + \mathbf{w}_2, t-2). \quad (7.12)$$

Ce faisant, sous les hypothèses de validité de (2.39), nous aurons bien construit une prédiction de $I(\mathbf{p}, t)$ au bruit près. Nous nous sommes ainsi ramenés au cas stationnaire et donc avons bien *compensé les mouvements transparents*.

7.2.2 Limitations de la compensation des mouvements transparents

Choix du gain. En considérant la relation (7.12), on peut écrire une relation du type (7.1) pour le débruitage comme suit :

$$\begin{aligned} \hat{I}(\mathbf{p}, t) &= \alpha(\mathbf{p}, t)I(\mathbf{p}, t) \\ &+ (1 - \alpha(\mathbf{p}, t))(I(\mathbf{p} + \mathbf{w}_1, t-1) \\ &\quad + I(\mathbf{p} + \mathbf{w}_2, t-1) - I(\mathbf{p} + \mathbf{w}_1 + \mathbf{w}_2, t-2)). \end{aligned} \quad (7.13)$$

Ce qui se traduit en terme de variance, sous l'hypothèse d'indépendance des bruits⁵ par :

$$\begin{aligned} \sigma_{\hat{I}}^2(t) &= \alpha(\mathbf{p}, t)^2 \sigma^2 \\ &+ (1 - \alpha(\mathbf{p}, t))^2 (\sigma_{\hat{I}}^2(\mathbf{p} + \mathbf{w}_1, t-1) \\ &\quad + \sigma_{\hat{I}}^2(\mathbf{p} + \mathbf{w}_2, t-1) + \sigma_{\hat{I}}^2(\mathbf{p} + \mathbf{w}_1 + \mathbf{w}_2, t-2)). \end{aligned} \quad (7.14)$$

Comme expliqué plus haut, puisqu'il est illusoire d'estimer l'écart-type de bruit très localement, nous abandonnons la dépendance vis à vis de la position spatiale :

$$\sigma_{\hat{I}}^2(t) = \alpha(t)^2 \sigma^2 + (1 - \alpha(t))^2 (2\sigma_{\hat{I}}^2(t-1) + \sigma_{\hat{I}}^2(t-2)). \quad (7.15)$$

Nous souhaitons choisir le gain $\alpha(t)$ minimisant le bruit de l'image traitée $\sigma_{\hat{I}}^2(t)$:

$$\frac{d\sigma_{\hat{I}}^2(t)}{d\alpha(t)} = 2\left(\alpha(t)\sigma^2 + (\alpha(t) - 1)(2\sigma_{\hat{I}}^2(t-1) + \sigma_{\hat{I}}^2(t-2))\right) = 0. \quad (7.16)$$

D'où finalement

$$\alpha(t) = \frac{2\sigma_{\hat{I}}^2(t-1) + \sigma_{\hat{I}}^2(t-2)}{2\sigma_{\hat{I}}^2(t-1) + \sigma_{\hat{I}}^2(t-2) + \sigma^2}. \quad (7.17)$$

⁵L'hypothèse d'indépendance spatiale des bruits n'est d'ailleurs pas exactement vérifiée, notamment à cause de la MTF et de problèmes d'interpolations si \mathbf{w}_1 et \mathbf{w}_2 sont proches.

Étude théorique de la limite de débruitage. Avec ce réglage (qui ne suppose notamment aucune adaptation du gain), (7.15) devient :

$$\sigma_{\hat{I}}^2(t) = \frac{(2\sigma_{\hat{I}}^2(t-1) + \sigma_{\hat{I}}^2(t-2))^2 \sigma^2 + \sigma^4 (2\sigma_{\hat{I}}^2(t-1) + \sigma_{\hat{I}}^2(t-2))}{(2\sigma_{\hat{I}}^2(t-1) + \sigma_{\hat{I}}^2(t-2) + \sigma^2)^2}. \quad (7.18)$$

Nous étudions en détail la convergence de cette suite dans l'annexe F. Nous montrons qu'elle converge vers son unique point fixe stable :

$$\lim_{t \rightarrow \infty} \sigma_{\hat{I}}^2(t) = \frac{2\sigma^2}{3}. \quad (7.19)$$

Nous sommes donc loin du niveau de débruitage asymptotique du filtre récursif compensé en mouvement sur les images vidéos classiques puisque nous nous heurtons à la limite haute de :

$$\lim_{t \rightarrow \infty} \sigma_{\hat{I}}(t) = \sqrt{\frac{2}{3}} \cdot \sigma \simeq 0.816\sigma. \quad (7.20)$$

Cette méthode permet certes de garder toute l'information pertinente, mais de n'ôter dans le meilleur des cas que 18,4% du bruit en écart-type.

Nous montrons d'autre part en annexe F que le filtrage compensé en mouvement transparent pour le cas de n couches mène à un débruitage asymptotique de $\sqrt{\frac{2^n-2}{2^n-1}}$ en écart-type, ce qui est très faible comme montré dans le tableau 7.4.

TAB. 7.4: Limite asymptotique de débruitage théorique en écart-type du filtre récursif temporel compensé en mouvement dans une situation à n couches.

n	1	2	3	4
$\sqrt{\frac{2^n-2}{2^n-1}}$	0	0,816	0,926	0.966

7.2.3 Résultats

Nous vérifions ces résultats théoriques sur la séquence synthétique déjà utilisée précédemment. Nous confrontons les niveaux de bruit théoriques à ceux que nous mesurons sur cette expérimentation dans le tableau 7.5. Nous observons qu'ils sont très proches, et donc notamment que l'utilisation d'interpolations ne modifie pas notablement le comportement asymptotique du filtre. Ces tests confirment que la réduction de la variance du bruit effective est faible.

Nous présentons également à la figure 7.4 la cinquième image traitée par le filtre temporel compensé en mouvements transparents. La comparaison avec le filtre adaptatif sans compensation de mouvement révèle un débruitage moins marqué, mais aussi une absence totale d'artefacts : tous les détails présents sur la séquence originale sont préservés. Le gain en terme de débruitage par rapport à la séquence originale, s'il est peu apparent sur les

TAB. 7.5: Evolution du gain et des niveaux de bruit théorique et pratique d'une séquence traitée par le filtre temporel récursif compensé en mouvements transparents.

t	2	3	4	5	6	7	8
α théorique	0.75	0.71	0.69	0.68	0.67	0.67	0.67
$\sigma_{\hat{f}}(t)/\sigma$ théorique	0.87	0.85	0.83	0.82	0.82	0.82	0.82
$\sigma_{\hat{f}}(t)/\sigma$ mesurée	0.87	0.82	0.79	0.79	0.78	0.78	0.78

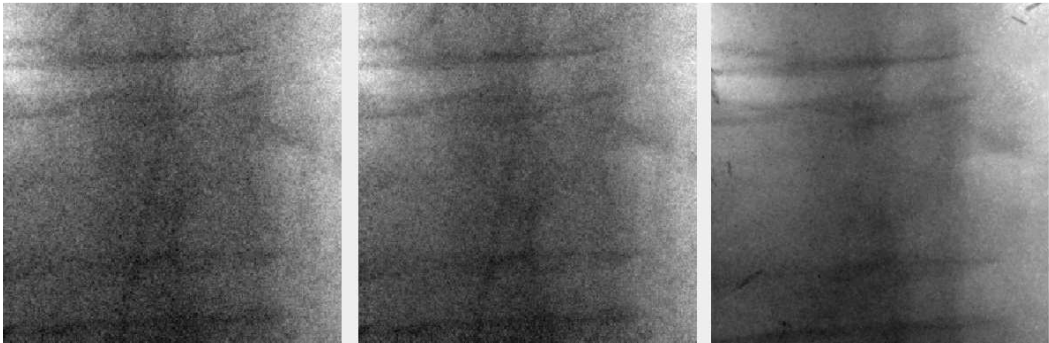


FIG. 7.4: Résultats du débruitage compensé en mouvements transparents d'une séquence synthétique dont les deux couches sont en mouvement. Nous présentons la cinquième image de la séquence débruitée au centre, et la comparons à gauche avec l'image bruitée initiale, et à droite avec la cinquième image de la séquence débruitée par le filtre temporel adaptatif sans compensation de mouvement.

images imprimées, est tout à fait appréciable lorsque l'on visualise les séquences d'images à $30Hz$.

D'autre part, cette approche est relativement robuste à des estimations imprécises des mouvements. Une erreur d'estimation de l'ordre du pixel ne sera pas dramatique car les anatomies que nous suivons sont pour la plupart nettement plus larges. En outre, le gain utilisé est toujours supérieur à $2/3$, de sorte que les pertes de contraste en cas d'estimation imprécise ne seront jamais très importantes.

7.3 Conclusion.

La méthode de débruitage compensée en mouvements transparents apporte une amélioration indiscutable car à peu près 20% du bruit est filtré sans qu'aucun artefact ne vienne modifier la séquence originale par ailleurs. En fait, sa capacité de débruitage limitée a l'avantage de contrôler les conséquences d'estimations imprécises concernant la perte de contraste.

Cependant, sa mise en œuvre suppose que les mouvements transparents aient été préalablement estimés. Comme nous l'avons vu dans les chapitres 2 à 5, il s'agit d'une tâche qui reste coûteuse. Il est donc difficilement envisageable industriellement d'investir dans

la résolution en temps réel de ce problème pour un apport encore restreint en terme de débruitage, même si aucun artefact n'est introduit. Nous proposons donc dans le chapitre suivant un filtre temporel permettant de débruiter de manière plus marquée les séquences d'images fluoroscopiques.

Chapitre 8

Débruitage temporel hybride de séquences d'images transparentes

Nous avons constaté dans le chapitre 7 les limites du filtre temporel récursif adaptatif : il pousse à un compromis entre conservation des contrastes et qualité du débruitage que nous aimerions dépasser. Nous avons proposé dans le chapitre précédent de compenser les mouvements transparents pour être en mesure de filtrer sans atténuer le contraste des objets.

Il se trouve que cette stratégie ne permet pas un débruitage poussé, puisqu'elle se heurte à une limite de débruitage restrictive (81,6% en écart-type). Le filtre résultant n'introduit donc pas d'artefacts, mais débruite relativement peu, ce qui est rédhibitoire étant donné sa complexité¹.

Nous proposons dans ce chapitre un filtre temporel dit *hybride*. Il cherche à ne compenser les mouvements que pour les pixels pour lesquels c'est nécessaire, ce qui permet de filtrer de manière plus marquée les autres. Nous présentons dans une première section le principe de l'approche, puis nous développons la méthode définie. Une troisième section détaille la façon dont la configuration "aucune, une ou deux couches homogènes" est détectée, en faisant un parallèle avec le filtrage adaptatif. Finalement, des résultats sur données synthétiques et réelles seront présentés.

8.1 Filtre temporel hybride

8.1.1 Principe

Le filtrage compensé en mouvements transparents permet de conserver les détails de l'image, mais souffre d'une capacité de débruitage limitée. Or, les images d'examens vasculaires que nous considérons sont caractérisées par de nombreuses zones homogènes peu contrastées qu'il est inutile de préserver exactement. Nous présentons dans cette section un filtre temporel *hybride*, qui :

- est *compensé en mouvements transparents* sur les régions où cela est nécessaire, c'est-à-dire aux pixels où les deux couches considérées sont hétérogènes. Cela permet de conserver l'information portée par ces pixels ;

¹Ou plus exactement la complexité de l'estimation des mouvements transparents sur laquelle il s'appuie.

- n'est pas compensé en mouvements transparents sur les régions où une couche au moins est homogène, ce qui permet d'y réduire le bruit de manière plus importante. Nous compensons alors seulement le mouvement de la couche la plus texturée avant de filtrer temporellement. En d'autres termes, nous effectuons dans ce cas une *compensation de mouvement non transparent*.

8.1.2 Quatre configurations locales différentes

Quatre configurations locales doivent être considérées en chaque pixel \mathbf{p} :

- **(C₀)** : les deux couches sont hétérogènes en \mathbf{p} : pour préserver les détails, il faut filtrer en compensant les mouvements transparents en s'appuyant sur $I(\mathbf{p}, t)$ et $\bar{I}(\mathbf{p}, t)$, où

$$\bar{I}(\mathbf{p}, t) = \hat{I}(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t-1) + \hat{I}(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t-1) - \hat{I}(\mathbf{p} + \mathbf{w}_1(\mathbf{p}) + \mathbf{w}_2(\mathbf{p}), t-2). \quad (8.1)$$

- **(C₁)** : la première couche² seulement est hétérogène dans le voisinage considéré. On peut alors se contenter de compenser les mouvements de la première couche seulement, car :

$$I(\mathbf{p}, t) = I_1(\mathbf{p}, t) + I_2(\mathbf{p}, t) \quad (8.2)$$

$$= I_1(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t-1) + I_2(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t-1). \quad (8.3)$$

La deuxième couche étant homogène, les valeurs de I_2 varient peu en fonction de la position, d'où :

$$I(\mathbf{p}, t) \simeq I_1(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t-1) + I_2(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t-1) \quad (8.4)$$

$$= I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t-1). \quad (8.5)$$

Dès lors, nous pouvons réduire le bruit en \mathbf{p} en nous appuyant sur $I(\mathbf{p}, t)$, $\hat{I}(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t-1)$ et $\bar{I}(\mathbf{p}, t)$, puisque l'on mesure dans ces trois cas l'intensité du même point physique (à un bruit près).

- **(C₂)** : De la même façon, si la deuxième couche seulement est hétérogène, on s'appuyera sur $I(\mathbf{p}, t)$, $\hat{I}(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t-1)$ et $\bar{I}(\mathbf{p}, t)$.
- **(C₃)** : Enfin, si les deux couches sont localement homogènes, quatre mesures seront exploitables pour débruiter \mathbf{p} : $I(\mathbf{p}, t)$, $\hat{I}(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t-1)$, $\hat{I}(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t-1)$ et $\bar{I}(\mathbf{p}, t)$

8.1.3 Une cinquième configuration

Pour être utilisable sur des séquences d'images réelles, sur lesquelles

- les mouvements ne sont pas nécessairement bien estimés,
- les outils interventionnels sont ignorés,

²Ce choix de première ou deuxième couche est évidemment arbitraire. Nous nommons un des deux modèles de mouvement θ_1 , et considérons la couche associée comme la couche numéro 1.

le filtre hybride doit savoir s'adapter au cas où les estimés des mouvements sont erronés. Nous introduisons pour cela une cinquième configuration portant sur la sortie de la méthode d'estimation du mouvement :

- (**C₄**) : Lorsque les estimés de mouvement sont erronés, nous nous contentons de conserver $I(\mathbf{p}, t + 1)$ comme valeur de l'image débruitée au point \mathbf{p} .

Les cinq configurations considérées sont schématisées à la figure 8.1, avec les stratégies de filtrage temporel adapté à chacun des cas. Notons que dans ce schéma, les quatre premières configurations supposent implicitement que les estimés des mouvements utilisés sont corrects.

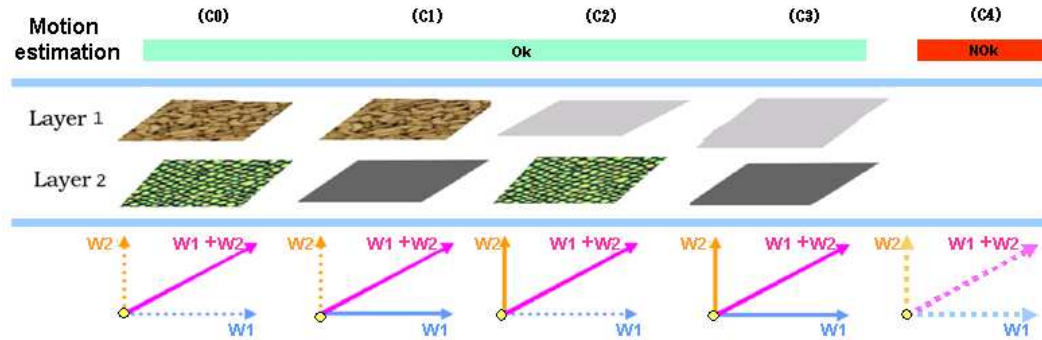


FIG. 8.1: Illustration des cinq configurations étudiées. De gauche à droite : C_0 (les deux couches sont hétérogènes), C_1 (la première couche seulement est hétérogène), C_2 (la deuxième couche est hétérogène), C_3 (les deux couches sont homogènes) et C_4 (les estimés des vitesses sont erronés). Le schéma dans la partie inférieure de la figure illustre quels types d'images (recalées par rapport à \mathbf{w}_1 seul, à \mathbf{w}_2 seul, aux deux mouvements transparents de manière combinée par le vecteur $\mathbf{w}_1 + \mathbf{w}_2$, ou à l'image acquise représentée par le point jaune à l'origine) sont exploitables par le filtre temporel.

8.2 Filtrage linéaire

Filtrage effectué. Nous expliquons plus loin comment nous décidons laquelle des cinq configurations prévaut au point \mathbf{p} . Une fois cette décision prise, nous appliquons au pixel considéré le filtre suivant :

$$\begin{aligned} \hat{I}(\mathbf{p}, t) = & \alpha(\mathbf{p})I(\mathbf{p}, t) + \beta(\mathbf{p})\hat{I}(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t - 1) \\ & + \delta(\mathbf{p})\hat{I}(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t - 1) + \gamma(\mathbf{p})\bar{I}(\mathbf{p}, t) \end{aligned} \quad (8.6)$$

avec bien sûr :

$$\begin{aligned} \bar{I}(\mathbf{p}, t) = & \hat{I}(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t - 1) \\ & + \hat{I}(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t - 1) - \hat{I}(\mathbf{p} + \mathbf{w}_1(\mathbf{p}) + \mathbf{w}_2(\mathbf{p}), t - 2) \end{aligned} \quad (8.7)$$

$$\text{et} \quad \alpha(\mathbf{p}) + \beta(\mathbf{p}) + \delta(\mathbf{p}) + \gamma(\mathbf{p}) = 1. \quad (8.8)$$

Dans l'hypothèse (**C₀**), $\beta(\mathbf{p})$ et $\delta(\mathbf{p})$ valent 0, alors que $\delta(\mathbf{p})$ vaut 0 dans (**C₁**), $\beta(\mathbf{p})$ vaut 0 dans (**C₂**) et $\beta(\mathbf{p})$, $\delta(\mathbf{p})$ et $\gamma(\mathbf{p})$ valent 0 dans (**C₄**).

Le filtre (8.6) est linéaire. Sous l'hypothèse de l'indépendance des bruits, nous pouvons déterminer les poids optimaux pour chacune des cinq situations (\mathbf{C}_i). Nous donnons les valeurs correspondantes de $\alpha(\mathbf{p})$, $\beta(\mathbf{p})$, $\delta(\mathbf{p})$ et $\gamma(\mathbf{p})$ dans le tableau 8.1. Les calculs associés sont détaillés dans l'annexe G.

Calcul du niveau de bruit. Comme dans le chapitre précédent, nous considérons que le bruit résiduel des images traitées ne dépend pas de la position spatiale considérée. Même si en toute rigueur le bruit est hétéroscédastique³ puisque les poids varient spatialement (comme nous le verrons plus bas), nous ne savons pas estimer de manière fiable son écart-type localement. En particulier, l'utilisation de la formule (8.6) pour estimer itérativement son niveau diverge très vite.

C'est pourquoi nous avons approché la valeur du bruit résiduel par la formule suivante :

$$\sigma_f(t) = \sigma(0,55 + 0,35e^{-\frac{t}{4}}) \text{ pour } t > 1 \quad (8.9)$$

établie sur la base de mesures effectuées sur une séquence simulée que nous avons traitée (Fig.8.2). Nous confrontons cette approximation avec les niveaux de bruit réels des images débruitées dans le tableau 8.2.

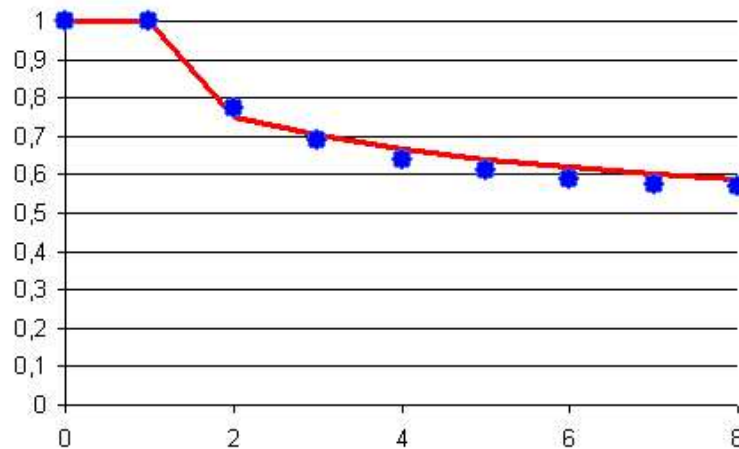


FIG. 8.2: Mesure du bruit résiduel sur une séquence d'images synthétiques traitée par le filtre hybride (tracé en bleu), modèle de bruit utilisé (tracé en rouge). Les ordonnées correspondent à l'écart-type du bruit normalisé ($\sigma_f(t)/\sigma$), et les abscisses à l'indice de l'image débruitée. Le filtre hybride utilisé est le FHPI (filtre hybride par patch itéré) décrit plus loin.

8.3 Détection de la configuration

8.3.1 Critère de décision.

Pour appliquer une de ces cinq stratégies, il nous faut d'abord décider si les couches sont homogènes ou non au pixel \mathbf{p} considéré. Discutons par exemple de l'homogénéité de

³Son écart-type varie spatialement.

TAB. 8.1: Réglages optimaux des gains dans les cinq configurations possibles. Nous notons σ l'écart-type du bruit de mesure, σ_f l'écart-type du bruit sur l'image débruitée précédente et $\bar{\sigma}$ l'écart-type du bruit de l'image prédite (dans ce document, cette notation note la prédiction et non la moyenne!).

Configuration	α	β
(C ₀)	$\frac{\bar{\sigma}^2}{\sigma^2 + \bar{\sigma}^2}$	0
(C ₁)	$\frac{\sigma_f^2 \bar{\sigma}^2}{\sigma_f^2 \bar{\sigma}^2 + \sigma^2 \bar{\sigma}^2 + \sigma_f^2 \sigma^2}$	$\frac{\sigma^2 \bar{\sigma}^2}{\sigma_f^2 \bar{\sigma}^2 + \sigma^2 \bar{\sigma}^2 + \sigma_f^2 \sigma^2}$
(C ₂)	$\frac{\sigma_f^2 \bar{\sigma}^2}{\sigma_f^2 \bar{\sigma}^2 + \sigma^2 \bar{\sigma}^2 + \sigma_f^2 \sigma^2}$	0
(C ₃)	$\frac{\sigma_f^4 \bar{\sigma}^2}{\sigma_f^4 \bar{\sigma}^2 + \sigma^2 \sigma_f^2 \bar{\sigma}^2 + \sigma_f^2 \sigma^2 \bar{\sigma}^2 + \sigma_f^4 \sigma^2}$	$\frac{\sigma^2 \sigma_f^2 \bar{\sigma}^2}{\sigma_f^4 \bar{\sigma}^2 + \sigma^2 \sigma_f^2 \bar{\sigma}^2 + \sigma_f^2 \sigma^2 \bar{\sigma}^2 + \sigma_f^4 \sigma^2}$
(C ₄)	1	0

Configuration	δ	γ
(C ₀)	0	$\frac{\sigma^2}{\sigma^2 + \bar{\sigma}^2}$
(C ₁)	0	$\frac{\sigma_f^2 \sigma^2}{\sigma_f^2 \sigma^2 + \sigma^2 \bar{\sigma}^2 + \sigma_f^2 \sigma^2}$
(C ₂)	$\frac{\sigma^2 \bar{\sigma}^2}{\sigma_f^2 \bar{\sigma}^2 + \sigma^2 \bar{\sigma}^2 + \sigma_f^2 \sigma^2}$	$\frac{\sigma_f^2 \sigma^2}{\sigma_f^2 \bar{\sigma}^2 + \sigma^2 \bar{\sigma}^2 + \sigma_f^2 \sigma^2}$
(C ₃)	$\frac{\sigma_f^2 \sigma^2 \bar{\sigma}^2}{\sigma_f^4 \bar{\sigma}^2 + \sigma^2 \sigma_f^2 \bar{\sigma}^2 + \sigma_f^2 \sigma^2 \bar{\sigma}^2 + \sigma_f^4 \sigma^2}$	$\frac{\sigma_f^4 \sigma^2}{\sigma_f^4 \bar{\sigma}^2 + \sigma^2 \sigma_f^2 \bar{\sigma}^2 + \sigma_f^2 \sigma^2 \bar{\sigma}^2 + \sigma_f^4 \sigma^2}$
(C ₄)	0	0

la deuxième couche (dont les vecteurs de vitesse sont donnés par $\mathbf{w}_2(\mathbf{p})$). Comme expliqué plus haut, nous voulons établir si

$$I_2(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t - 1) \simeq I_2(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t - 1) \quad (8.10)$$

ce qui est équivalent d'après (8.5) à :

$$I(\mathbf{p}, t) \simeq I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t - 1). \quad (8.11)$$

Il s'agit donc de comparer $I(\mathbf{p}, t)$ et $I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t - 1)$, et de décider si ces deux valeurs sont "égales". En pratique, il s'agit de déterminer si elles diffèrent seulement à cause du bruit (la couche est homogène), ou bien si elles correspondent à des points physiques différents (la couche est hétérogène).

Cette problématique correspond exactement à celle du filtrage adaptatif, quand on cherche à déterminer si l'on considère bien deux pixels d'une région stationnaire.

Il faut également décider si les estimés du mouvement au pixel considéré sont fiables. Nous nous basons pour ce faire sur la comparaison de $I(\mathbf{p}, t)$ et de $\tilde{I}(\mathbf{p}, t)$, ce qui correspond également au problème considéré par les filtres adaptatifs.

Nous appliquons donc le même traitement que celui qui est effectué par les filtres adaptatifs. Plutôt que de décider de manière binaire de l'homogénéité de la couche et de la qualité de l'estimation des mouvements,

- nous calculons en chaque pixel \mathbf{p} deux facteurs correspondant à la configuration locale des deux couches : $h_1(\mathbf{p})$ qui représente le facteur associé à l'hypothèse "la couche 1 est homogène" (par l'équation (8.15) ci-dessous) et $h_2(\mathbf{p})$ qui représente le facteur associé à l'hypothèse "la couche 2 est homogène" (par (8.16)). Ces deux facteurs sont normalisés entre 0 et 1.

Comme les gains d'adaptation "classiques", ce sont des fonctions décroissantes de $|I(\mathbf{p}, t) - I(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t - 1)|$ (resp. $|I(\mathbf{p}, t) - I(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t - 1)|$).

De la même manière, un facteur $h_{12}(\mathbf{p})$ est associé à l'hypothèse " $\tilde{I}(\mathbf{p}, t)$ est une bonne prédiction de $I(\mathbf{p}, t)$ ". Il s'agit d'une fonction décroissante de $|I(\mathbf{p}, t) - \tilde{I}(\mathbf{p}, t)|$.

Ainsi,

- la configuration (\mathbf{C}_0) (les estimés des mouvements sont fiables et les deux couches sont hétérogènes) se voit associée le facteur $h_{12}(1 - h_1)(1 - h_2)$,
 - la configuration (\mathbf{C}_1) (estimés de mouvements fiables, couche 1 hétérogène et couche 2 homogène), le facteur $h_{12}(1 - h_1)h_2$,
 - la configuration (\mathbf{C}_2) (estimés de mouvements fiables, couche 1 homogène et couche 2 hétérogène), le facteur $h_{12}(1 - h_1)h_1(1 - h_2)$,
 - la configuration (\mathbf{C}_3) (estimés de mouvements fiables, deux couches homogènes) le facteur $h_{12}h_1h_2$,
 - et la configuration (\mathbf{C}_4) (estimés de mouvements erronés) $1 - h_{12}$.
- Nous filtrons la séquence d'images dans chacune des hypothèses (\mathbf{C}_i), et obtenons cinq images $\hat{I}_{(\mathbf{C}_i)}(\mathbf{p}, t)$.

– Nous combinons ces cinq images à la manière des filtres adaptatifs⁴ :

$$\begin{aligned}\hat{I}(\mathbf{p}, t) &= h_{12}(\mathbf{p})(1 - h_1(\mathbf{p}))(1 - h_2(\mathbf{p}))\hat{I}_{(\mathbf{C}_0)}(\mathbf{p}, t) \\ &\quad + h_{12}(\mathbf{p})(1 - h_1(\mathbf{p}))h_2(\mathbf{p})\hat{I}_{(\mathbf{C}_1)}(\mathbf{p}, t) \\ &\quad + h_{12}(\mathbf{p})h_1(\mathbf{p})(1 - h_2(\mathbf{p}))\hat{I}_{(\mathbf{C}_2)}(\mathbf{p}, t) \\ &\quad + h_{12}(\mathbf{p})h_1(\mathbf{p})h_2(\mathbf{p})\hat{I}_{(\mathbf{C}_3)}(\mathbf{p}, t) \\ &\quad + (1 - h_{12}(\mathbf{p}))\hat{I}_{(\mathbf{C}_4)}(\mathbf{p}, t)\end{aligned}\tag{8.13}$$

avec

$$\begin{aligned}\hat{I}_{(\mathbf{C}_i)}(\mathbf{p}, t) &= \alpha_{(\mathbf{C}_i)}(\mathbf{p})I(\mathbf{p}, t) + \beta_{(\mathbf{C}_i)}(\mathbf{p})\hat{I}(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t - 1) \\ &\quad + \delta_{(\mathbf{C}_i)}(\mathbf{p})\hat{I}(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t - 1) + \gamma_{(\mathbf{C}_i)}(\mathbf{p})\bar{I}(\mathbf{p}, t).\end{aligned}\tag{8.14}$$

La figure 8.4 illustre le fonctionnement du filtre hybride que nous venons de présenter.

8.3.2 Fonction de décision

Spécification des facteurs h . Comme dans le cas du filtre adaptatif, le facteur h_1 (resp. h_2 , h_{12}) est une fonction décroissante de $|I(\mathbf{p}, t) - \hat{I}(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t - 1)|$ (resp. $|I(\mathbf{p}, t) - \hat{I}(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t - 1)|$, $|I(\mathbf{p}, t) - \tilde{I}(\mathbf{p}, t)|$) :

$$h_1(\mathbf{p}) = \psi(|I(\mathbf{p}, t) - \hat{I}(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t - 1)|)\tag{8.15}$$

$$h_2(\mathbf{p}) = \psi(|I(\mathbf{p}, t) - \hat{I}(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t - 1)|).\tag{8.16}$$

$$h_{12}(\mathbf{p}) = \psi(|I(\mathbf{p}, t) - \tilde{I}(\mathbf{p}, t)|).\tag{8.17}$$

L'expression exacte de $\psi(\cdot)$ doit être déterminée de façon heuristique en fonction du problème posé. Nous choisissons une fonction en trois parties du même type que la courbe rouge de la figure 7.1 (Fig.8.3). Elle permet de distinguer des régions où l'hypothèse considérée est très probable (et se voient associer la valeur 1) de régions où elle est très peu probable (pour lequel le facteur est choisi presque nul⁵). Ces deux intervalles sont séparés par une zone permettant une transition douce entre les deux types de décision.

La forme du facteur est donc définie par deux seuils. Nous avons en pratique opté pour un premier seuil à $s_1 = 1.5\sigma$ et un deuxième seuil à $s_2 = 2\sigma$, où σ représente le bruit d'acquisition (mesuré par le système d'imagerie).

⁴En pratique, nous calculons directement (8.13) sans construire explicitement les images intermédiaires $\hat{I}_{(\mathbf{C}_i)}(\mathbf{p}, t)$. Nous combinons les cinq configurations au niveau du calcul des poids $\alpha(\mathbf{p})$, $\beta(\mathbf{p})$, $\delta(\mathbf{p})$ et $\gamma(\mathbf{p})$, en formant :

$$\begin{aligned}\alpha(\mathbf{p}) &= h_{12}(\mathbf{p})(1 - h_1(\mathbf{p}))(1 - h_2(\mathbf{p}))\alpha_{(\mathbf{C}_0)}(\mathbf{p}) + h_{12}(\mathbf{p})(1 - h_1(\mathbf{p}))h_2(\mathbf{p})\alpha_{(\mathbf{C}_1)}(\mathbf{p}) \\ &\quad + h_{12}(\mathbf{p})h_1(\mathbf{p})(1 - h_2(\mathbf{p}))\alpha_{(\mathbf{C}_2)}(\mathbf{p}) + h_{12}(\mathbf{p})h_1(\mathbf{p})h_2(\mathbf{p})\alpha_{(\mathbf{C}_3)}(\mathbf{p}) \\ &\quad + (1 - h_{12}(\mathbf{p}))\alpha_{(\mathbf{C}_4)}(\mathbf{p})\end{aligned}\tag{8.12}$$

où $\alpha_{(\mathbf{C}_i)}(\mathbf{p})$ est le poids $\alpha(\mathbf{p})$ utilisé dans l'hypothèse (\mathbf{C}_i) (Tab.8.1). Les poids $\beta(\mathbf{p})$, $\delta(\mathbf{p})$ et $\gamma(\mathbf{p})$ sont calculés de la même façon, et l'image finale est obtenue par (8.6).

⁵Il n'est pas exactement nul pour se prévenir contre le bruit "poivre et sel".

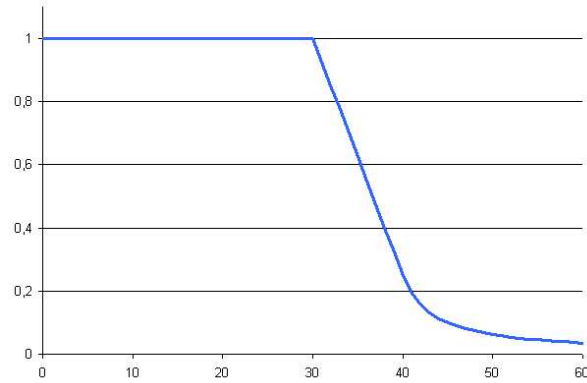


FIG. 8.3: Facteur h_1 (resp h_2, h_{12}) du filtre temporel hybride en fonction de $|I(\mathbf{p}, t) - \hat{I}(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t)|$ (resp $|I(\mathbf{p}, t) - \hat{I}(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t)|, |I(\mathbf{p}, t) - \tilde{I}(\mathbf{p}, t)|$).

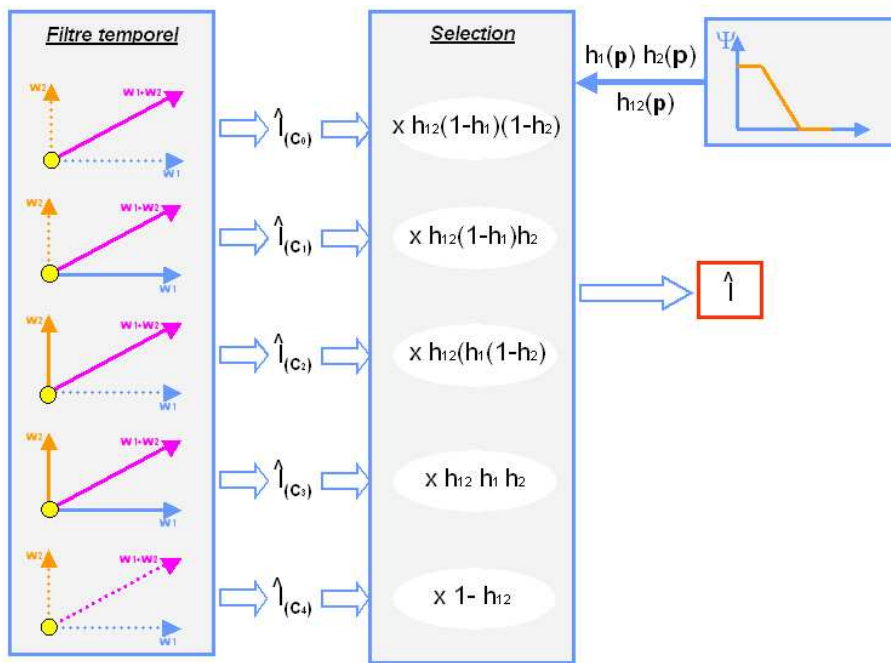


FIG. 8.4: Schéma de fonctionnement de filtre hybride. Il est composé de deux modules : le premier filtre temporellement les images dans les cinq configurations possibles, le second pondère les cinq images intermédiaires selon la “probabilité” de leur occurrence au pixel considéré.

Prétraitement des images. Les facteurs spécifiés seront d’autant plus fiables que les images sur lesquelles nous nous basons seront peu bruitées. Nous pouvons donc envisager différents prétraitements pour $I(\mathbf{p}, t)$ (et éventuellement pour $\hat{I}(\mathbf{p} + \mathbf{w}_i(\mathbf{p}), t - 1)$ ($i = 1, 2$) et $\tilde{I}(\mathbf{p}, t)$) afin d’accroître la pertinence des facteurs h_1, h_2 et h_{12} . Nous notons le filtre

hybride sans prétraitement *FHD* (Filtre Hybride Direct).

Insistons sur le fait que ce prétraitement est interne au module de sélection de la configuration : les images prétraitées n'entrent en jeu que dans la détermination des facteurs h_1 , h_2 et h_{12} , mais aucunement dans le filtrage proprement dit défini par la relation (8.6). Ceci permet d'atténuer considérablement l'impact du prétraitement sur l'esthétique de bruit de l'image finale⁶.

Nous proposons trois prétraitements :

- *Sélection itérée*. C'est le filtre hybride lui-même qui sert de prétraitement.

La méthode de débruitage est donc appliquée dans un premier temps pour produire une image $\hat{I}^{(1)}(\mathbf{p}, t)$. Dans un deuxième temps, le filtre hybride est réappliqué aux images initiales, mais on substitue $I(\mathbf{p}, t)$ par $\hat{I}^{(1)}(\mathbf{p}, t)$ dans le module de sélection. Il en résulte l'image traitée finale.

Nous notons le filtre comprenant ce prétraitement *FHDI* (FHD Itéré).

- *Préfiltrage spatial (éventuellement itéré)*. Nous proposons alternativement de traiter $\hat{I}(\mathbf{p} + \mathbf{w}_i(\mathbf{p}), t - 1)$ ($i = 1, 2$) et $I(\mathbf{p}, t)$ par un filtre spatial adaptatif avant le calcul des facteurs h . Nous avons choisi une variation du filtre bilatéral [Tomasi 98] (sans terme dépendant de la distance spatiale) :

$$\hat{I}^{(2)}(\mathbf{p}, t) = \frac{1}{\sum_{\mathbf{p}' \in V(\mathbf{p})} w(\mathbf{p}, \mathbf{p}', t)} \sum_{\mathbf{p}' \in V(\mathbf{p})} w(\mathbf{p}, \mathbf{p}', t) I(\mathbf{p}', t) \quad (8.18)$$

$$\text{avec } w(\mathbf{p}, \mathbf{p}', t) = \exp\left(-\frac{(I(\mathbf{p}, t) - I(\mathbf{p}', t))^2}{(2Z\sigma(\mathbf{p}, t))^2}\right). \quad (8.19)$$

Nous utilisons un filtre "bilatéral" relativement conservateur ($Z = 3$ sur un voisinage 3×3), car nous ne voulons surtout pas gommer de détails à ce stade. Nous gardons en outre la possibilité d'itérer la sélection de la configuration comme plus haut, et choisissons pour seuils $0,75\sigma$ et $1,5\sigma$ au premier passage, puis $0,5\sigma$ et σ au deuxième passage éventuel. Le filtre correspondant sera dénommé *FHB* (Filtre Hybride avec prétraitement par le filtre Bilatéral), et *FHBI* s'il est itéré.

- *Utilisation de patch (éventuellement itéré)*. Enfin, nous proposons une troisième alternative qui consiste non pas à sélectionner la configuration sur la seule valeur de $|I(\mathbf{p}, t) - \hat{I}(\mathbf{p} + \mathbf{w}_i(\mathbf{p}), t - 1)|$ pour décider de l'homogénéité de la couche considérée en \mathbf{p} , mais sur l'expression

$$\sum_{\mathbf{p}' \in V(\mathbf{p})} |I(\mathbf{p}', t) - \hat{I}(\mathbf{p} + \mathbf{w}_i(\mathbf{p}'), t - 1)|, \quad i = 1, 2 \quad (8.20)$$

définie sur un petit voisinage $V(\mathbf{p})$ (en pratique 3×3) de \mathbf{p} .

Ce prétraitement peut également être utilisé dans une sélection de configuration itérée : un premier filtre hybride avec prétraitement par motif (ou "patch" en anglais) est utilisé afin de construire une image débruitée qui sera utilisée dans le module de sélection de configuration d'un nouveau filtre hybride par motif. Les seuils valent $0,75\sigma$ et $1,5\sigma$ au premier passage, puis $0,5\sigma$ et σ au deuxième. Ce filtre sera noté *FHP* (Filtre Hybride par Patch), et *FHPI* s'il est itéré.

⁶Il aura tout de même un impact indirect, car un filtre de prétraitement spatial, en corrélant le bruit, produira des cartes des facteurs h_1 , h_2 et h_{12} légèrement corrélées spatialement. D'où une stratégie de filtrage dont la légère corrélation spatiale se devinera sur l'image finale. Mais ceci est du second ordre.

TAB. 8.2: Evolution du niveau de bruit résiduel normalisé $\hat{\sigma}(\mathbf{p}, t)/\sigma$ d'une séquence réaliste débruitée par différentes versions du filtre hybride (les mouvements des deux couches sont connus).

t	2	3	4	5	6	7	8
Valeur prédite par (8.9)	0.71	0.66	0.63	0.60	0.58	0.56	0.55
FHD (direct)	0.77	0.69	0.65	0.62	0.61	0.60	0.59
FHB (bilatéral)	0.72	0.66	0.62	0.60	0.59	0.58	0.57
FHP (patch)	0.75	0.68	0.64	0.61	0.60	0.59	0.58
FHDI	0.76	0.68	0.63	0.60	0.59	0.58	0.58
FHBI	0.72	0.66	0.61	0.59	0.57	0.56	0.55
FHPI	0.76	0.68	0.64	0.61	0.60	0.58	0.57

Récapitulons les acronymes introduits pour désigner les différentes versions du filtre hybride :

- *FHD* : Filtre Hybride Direct (pas de prétraitement).
- *FHI* : Filtre Hybride avec sélection Itérée.
- *FHB* et *FHBI* : Filtre Hybride avec prétraitement Bilatéral (éventuellement Itéré).
- *FHP* et *FHPI* : Filtre Hybride avec utilisation de Patch (éventuellement Itéré).

8.4 Résultats à mouvements connus

8.4.1 Traitement de la séquence test

Nous avons appliqué ces différentes versions du filtre hybride à la séquence fluoroscopique réaliste sur laquelle nous avons testé les autres algorithmes. Les résultats sont donnés dans le tableau 8.2, pour le cas où les mouvements sont connus. Nous utilisons les valeurs vraies des paramètres des modèles de mouvement (ie, ceux qui ont été utilisés dans la génération des séquences d'images) dans les différents filtres testés.

Nous avons divisé le tableau en deux parties selon que la sélection est itérée après débruitage temporel ou non. Les performances en terme de débruitage sont relativement proche. Nous ne gardons alors que trois points de fonctionnement de l'algorithme : le filtre hybride direct FHD (réduction de bruit de 41%, $PSNR = 26, 7$), le filtre FHBI (réduction de bruit de 45%, $PSNR = 27, 2$) et le filtre FHPI (réduction de bruit de 43%, $PSNR = 27, 1$). Ceci est à rapprocher des 30% de réduction de bruit apporté par le filtre adaptatif non compensé FTA, et les 18,4% de réduction maximale théorique du filtre compensé en mouvement du chapitre précédent, sur la même séquence.

Il ne faut toutefois pas se contenter d'évaluer ces versions du filtre hybride d'après la seule métrique liée à la diminution de l'écart-type. La conservation des contrastes et l'esthétique de l'image résultante ont aussi une grande importance.

Pour juger de manière quantitative de la conservation de l'information, nous reproduisons le test précédent un grand nombre de fois (en ne faisant varier que les seules réalisations du bruit) et moyennons les séquences d'images obtenues. Ce faisant, l'effet du bruit disparaîtra et l'écart à la vérité terrain s'expliquera seulement par les distorsions subies par l'image

Algorithme 5 Algorithme du filtre hybride adaptatif.

En chaque point (\mathbf{p}, t) ,

- Évaluation du facteur $h_1(\mathbf{p})$ associé à l'hypothèse "la couche 1 est homogène au pixel \mathbf{p} ", et $h_2(\mathbf{p})$ associé à l'hypothèse "la couche 2 est homogène au pixel \mathbf{p} ".

1. Préfiltrage de l'image $I(\mathbf{p}, t)$.
2. Calcul de $h_1(\mathbf{p}, t)$ et $h_2(\mathbf{p}, t)$ par la fonction de la figure 8.3.

- Évaluation du facteur $h_{12}(\mathbf{p})$ associé à l'hypothèse "les mouvements estimés sont fiables au pixel \mathbf{p} ".

1. Préfiltrage de l'image $I(\mathbf{p}, t)$.
2. Calcul de $h_{12}(\mathbf{p}, t)$ par la fonction de la figure 8.3.

- Calcul des poids $\alpha_{(C_i)}(\mathbf{p})$, $\beta_{(C_i)}(\mathbf{p})$, $\delta_{(C_i)}(\mathbf{p})$ et $\gamma_{(C_i)}(\mathbf{p})$ dans chacune des cinq configurations.

Leur expression est donnée dans le tableau 8.1 en fonction du niveau de bruit des différentes images. Le bruit d'acquisition est connu, et le niveau de bruit des images traitées est approché par la formule (8.2).

- Calcul des poids du filtre hybride à partir des $\alpha_{(C_i)}(\mathbf{p})$ et $h_i(\mathbf{p}, t)$.

Par exemple,

$$\begin{aligned} \alpha_{id}(\mathbf{p}) = & h_{12}(\mathbf{p})(1 - h_1(\mathbf{p}))(1 - h_2(\mathbf{p}))\alpha_{(C_0)}(\mathbf{p}) + h_{12}(\mathbf{p})(1 - h_1(\mathbf{p}))h_2(\mathbf{p})\alpha_{(C_1)}(\mathbf{p}) \\ & + h_{12}(\mathbf{p})h_1(\mathbf{p})(1 - h_2(\mathbf{p}))\alpha_{(C_2)}(\mathbf{p}) + h_{12}(\mathbf{p})h_1(\mathbf{p})h_2(\mathbf{p})\alpha_{(C_3)}(\mathbf{p}) \\ & + (1 - h_{12}(\mathbf{p}))\alpha_{(C_4)}(\mathbf{p}) \end{aligned} \quad (8.21)$$

- Filtrage par la formule (8.6).
-

TAB. 8.3: Écart quadratique moyen entre la vérité terrain et la moyenne de 25 images débruitées pour différentes méthodes. Ceci permet de quantifier la qualité de la conservation de l'information pertinente. Les images considérées sont entachées d'un bruit typique d'images fluoroscopiques, d'écart-type $\sigma = 20$.

FTA	FHD	FHPI	FHBI
6.27	5.76	5.36	5.41

idéale [Aach 00]. Nous avons effectué cette expérience avec 25 générations d'images. Les résultats rapportés dans le tableau 8.3 ont trait à la huitième image de la séquence considérée (Fig.8.5). Ils révèlent une bien meilleure conservation de l'information par les filtres hybrides. Le FTA se traduit par une distorsion de 6.27 niveaux de gris (pour un bruit de mesure d'écart-type $\sigma = 20$), ce qui équivaut à un PSNR de 32.2 dB. Le FHD implique une distorsion de seulement 5.76 niveaux de gris, soit un PSNR de 32.9 dB, alors que le FHPI et le FHBI ont des performances supérieures équivalentes : 5.4 niveaux de gris, soit un PSNR de 33.5 dB. Ce gain de 1.3 dB en PSNR sur l'information préservée est extrêmement perceptible.

Cela dit, cette métrique ne rend pas compte du comportement spatial du filtre. Nous produisons pour en juger la huitième image des séquence débruitées en question sur la figure 8.5. Nous présentons également le bruit résiduel à la figure 8.6 (les différences des images débruitées avec l'image non bruitée de référence), pour mettre en évidence d'éventuelles pertes de contrastes locaux ou des artefacts de coloration du bruit.

L'élément le plus révélateur sur ces images est le cathéter qui traverse le centre de l'image de haut en bas. Il est difficile à détecter car il est peu contrasté, de sorte que le filtre adaptatif l'atténue notablement. Les filtres hybrides l'effacent moins car ils sont en mesure de suivre la couche correspondante. Seulement, plus ils sont réglés pour filtrer de manière relativement marquée, plus ils l'atténuent : le FHBI l'atténue notablement plus que le FHPI. Il n'y a par contre quasiment pas d'atténuation de l'anatomie dans les trois cas de filtres hybrides. Le bruit est très nettement réduit, ce qui est particulièrement notable quand on les compare avec le filtre compensé en mouvements transparents.

S'il n'y a que le filtre hybride direct FHD qui conserve l'esthétique originale du bruit, les deux autres filtres (FHBI et FHPI) produisent un bruit légèrement corrélé, mais qui reste tout à fait acceptable visuellement. L'introduction des différentes configurations entraîne certes une corrélation spatiale, mais elle est assez faible car le filtrage lui-même est purement temporel.

8.4.2 Étude des cartes de poids

Le filtre hybride produit des résultats plus intéressants que le filtre adaptatif (FTA) ou le filtre compensé en mouvements transparents. Pour mieux comprendre son fonctionnement, nous présentons à la figure 8.7 les cartes des valeurs des différents poids α , β , δ et γ pour la huitième image de la séquence.

Comme la majorité de l'image correspond à des zones homogènes, les quatre poids y sont le plus souvent non nuls. Les images mesurées $I(., t)$ et prédites $\tilde{I}(., t)$ étant entachées d'un

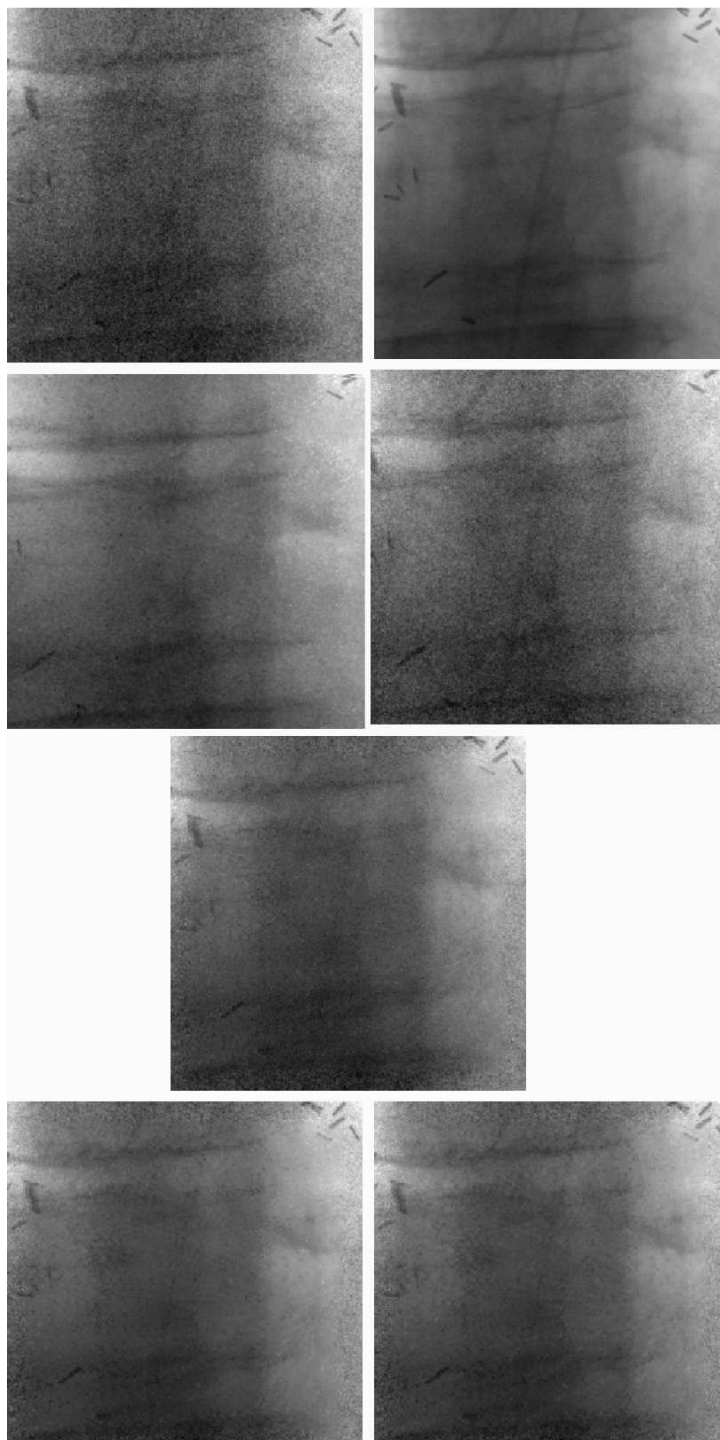


FIG. 8.5: Huitième image de la séquence synthétique traitée par différents filtres réducteurs de bruit. *En haut* à gauche l'image bruitée originale, à droite l'image idéale non bruitée correspondante. *Sur la deuxième ligne*, à gauche le résultat du filtrage temporel adaptatif (FTA) et à droite celui du filtrage temporel compensé en mouvements transparents. *Sur la troisième ligne*, le résultat du filtre temporel hybride direct (FHD). *Sur la dernière ligne*, à gauche le résultat du filtre temporel hybride "bilatéral itéré" (FHBI) et à droite celui du filtre hybride "patch itéré" (FHPI).

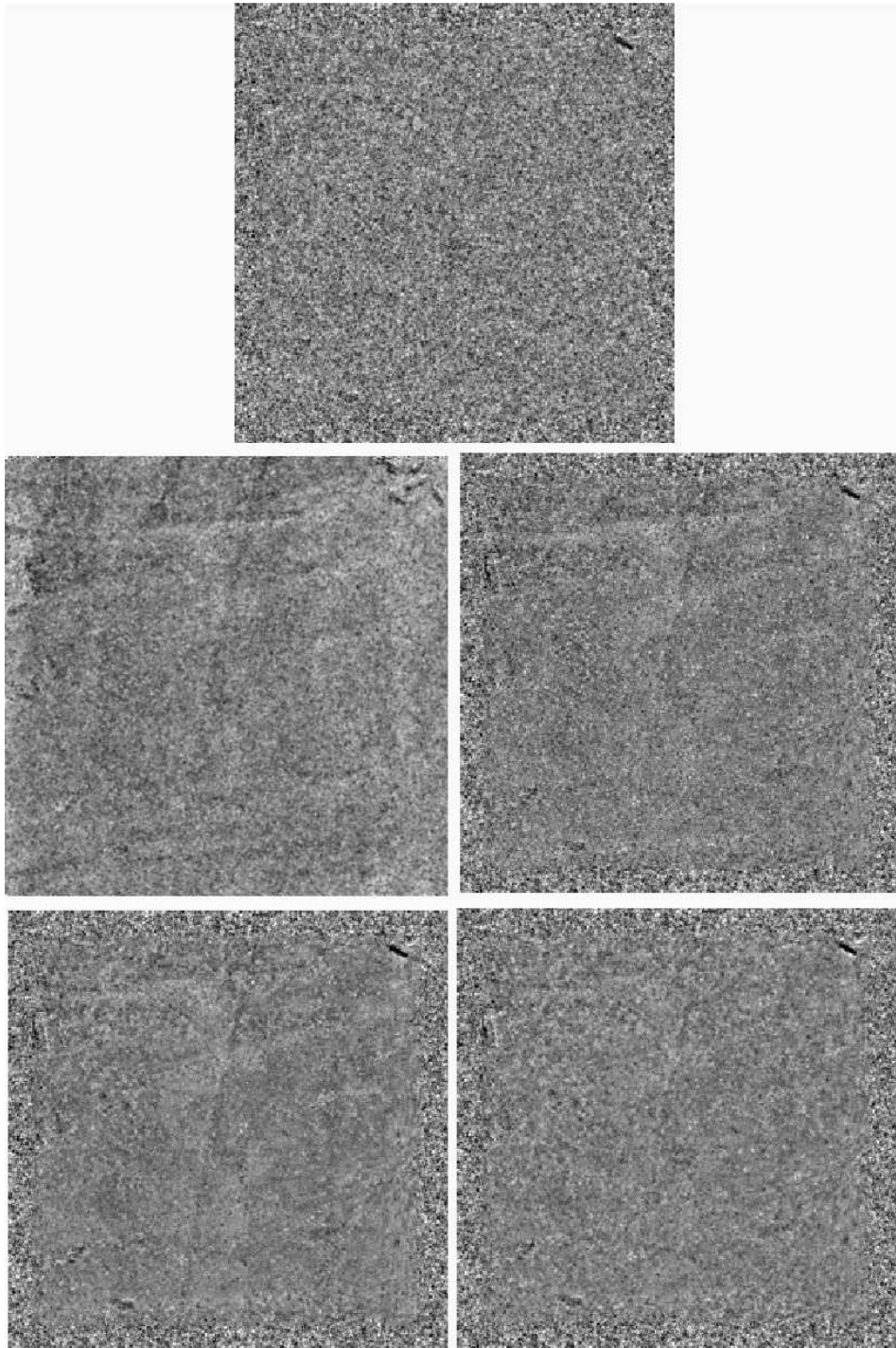


FIG. 8.6: Carte de bruit résiduel pour la huitième image de la séquence (différence entre la huitième image de la séquence débruitée et l'image idéale correspondante). *En haut* pour le cas du filtre récursif compensé en mouvement, *au milieu* à gauche pour un filtrage temporel récursif adaptatif (FTA), et à droite pour un filtre hybride direct (FHD), *en bas* pour un filtre hybride "bilatéral itéré" (FHBI) à gauche et pour un filtre hybride "patch itéré" (FHPI) en bas à droite. Les bords de l'image ne sont pas filtrés.

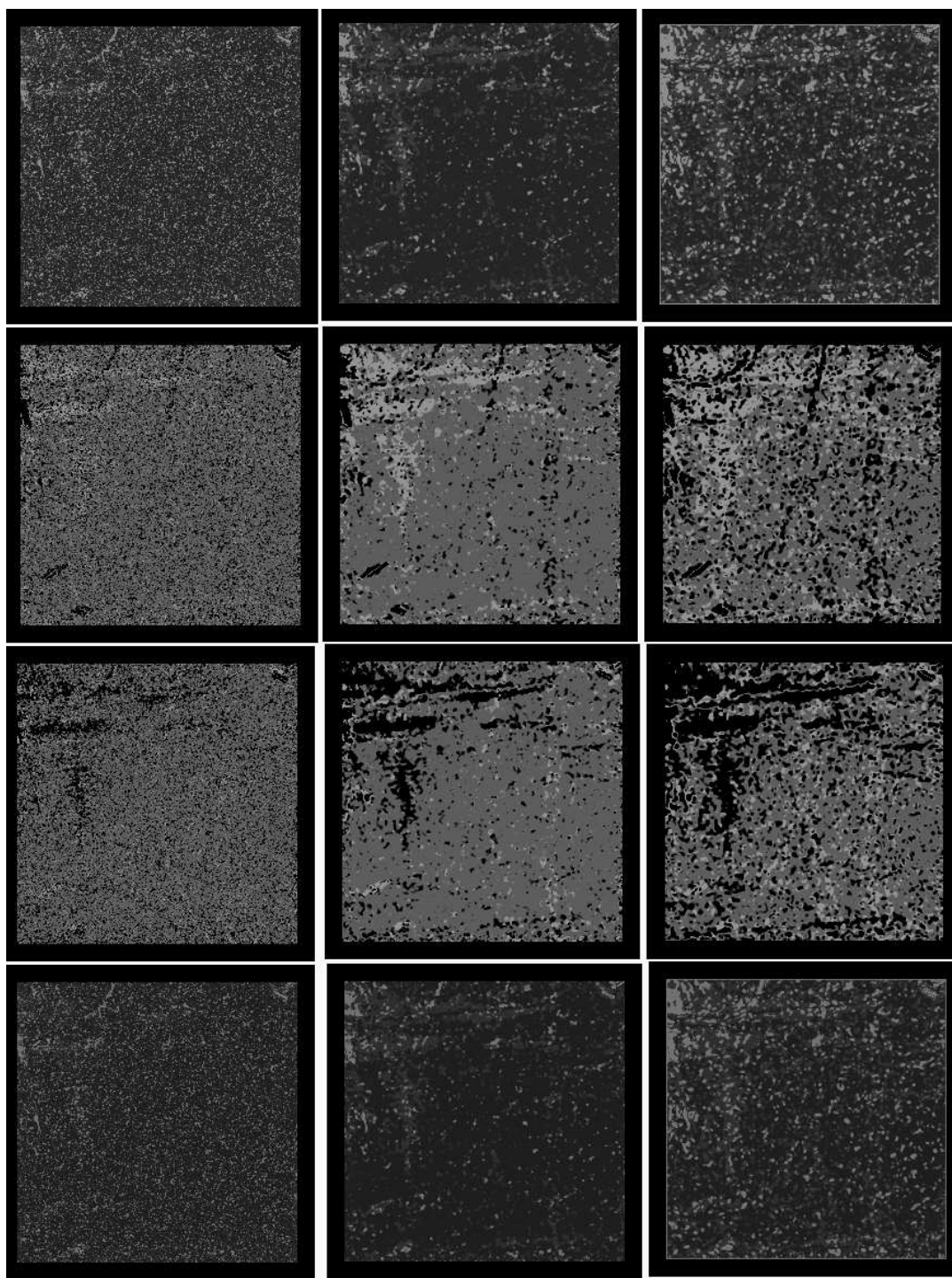


FIG. 8.7: Carte des poids pour les trois versions du filtre hybride temporel retenu : à gauche le filtre hybride direct, au centre le filtre hybride “bilatéral itéré” et à droite le filtre hybride “par patches” itéré. Pour chacun de ces trois filtres, nous présentons de haut en bas les cartes de α , β , δ et γ . Un niveau de gris proche du noir correspond à un poids nul, et un niveau de gris proche du blanc correspond à un poids de 1.

bruit supérieur à celui des images débruitées $\hat{I}(\cdot, t-1)$, on s'appuie plus fortement sur ces dernières ($\hat{I}(\cdot, t-1)$). Ceci explique que la majorité des valeurs de α et γ soient faibles, alors que celles de β et δ sont du même ordre (car les *deux* couches sont homogènes).

Les structures en mouvement sont le plus souvent détectées, qu'il s'agisse du cathéter, des agrafes ou de l'anatomie. On constate bien que seul le mouvement de la couche texturée est compensé, ce qui se traduit par un coefficient β (ou δ) nul. Des trois images restantes, c'est $\hat{I}(\cdot, t-1)$ qui est la moins bruitée, et se voit donc associer le poids le plus élevé.

Les cas où les *deux* couches sont texturées sont rares, de sorte que β et δ sont rarement nuls ensemble, sauf pour les pixels qui font leur apparition dans l'image à l'instant considéré.

Enfin, notons que les trois versions étudiées (FHD, FHBI, FHPI) donnent lieu à des cartes ayant des allures différentes. Les opérations spatiales incluses dans les approches FHBI et FHPI se traduisent par des cartes au grain plus grossier que l'approche FHD. Dans ce dernier cas, les décisions semblent par contre prises indépendamment en chaque pixel, ce qui rend notamment plus délicate la détection des structures d'intérêt.

La comparaison des cartes des poids dans les cas FHBI et FHPI met en évidence que les structures d'intérêt sont mieux détectées par la méthode "par patch", mais que les zones homogènes sont plus fortement filtrées par le filtre bilatéral, ce qui est cohérent avec les images de la Fig.8.6.

En conclusion, les trois variantes du filtrage hybride temporel correspondent aux points de fonctionnement suivants :

- le filtre hybride direct FHD est rapide, et il est celui qui conserve le mieux l'esthétique de l'image ;
- le filtre FHBI est le plus performant en ce qui concerne la réduction du bruit, mais aussi celui qui atténue le plus les contrastes ;
- le filtre FHPI permet un débruitage légèrement moins efficace, mais une meilleure préservation des contrastes.

Selon l'application finale visée et les contraintes en terme de temps de calcul, on peut donc arbitrer entre ces trois versions du filtre hybride. En tout état de cause, ils se comportent dans ces conditions idéales (le mouvement est connu) mieux que notre référence (le filtre adaptatif non compensé en mouvement), que ce soit en ce qui concerne la conservation des contrastes ou l'efficacité de la réduction de bruit.

8.5 Résultats avec estimation des mouvements transparents

Nous comparons dans cette section les performances en situation réelle (c'est-à-dire sans connaissance *a priori* des mouvements) des deux types de filtres temporels compensés en mouvement les plus compétitifs :

- le **filtre adaptatif temporel**, compensé en mouvement mono-couche. Nous le réglons comme en 7.1.3. Nous testons deux variantes de ce filtre, qui diffèrent par la méthode d'estimation mono-mouvement utilisée. Une première version s'appuie sur une mise en correspondance par blocs dense (un déplacement est calculé sur un bloc 32×32 pour chaque pixel de l'image). Nous la nommons par l'acronyme *FTACMD* (Filtre Temporel Adaptatif Compensé en mouvement Mono-couche, estimés par mise

TAB. 8.4: Evolution du niveau de bruit résiduel normalisé $\hat{\sigma}(\mathbf{p}, t)/\sigma$ d'une séquence synthétique réaliste traitée par un filtre adaptatif compensé en mouvement mono-couche (FTACM), ou un filtre temporel hybride (FHPI), avec estimation des mouvements. Les mouvements sont estimés, dans le cas mono-couche, soit par une mise en correspondance par blocs dense (D), soit par l'approche paramétrique (P) développée dans le chapitre 5. Dans le cas bi-couche, nous faisons appel à la méthode d'estimation en situation de bi-transparence simple (chapitre 3 -BS), ou en situation de bi-transparence distribuée (chapitre 5 - BD).

Indice de l'image	2	3	4	5	6	7	8
FTACMD	0,75	0,74	0,72	0,69	0,68	0,67	0,66
FTACMP	0,74	0,73	0,69	0,68	0,67	0,66	0,65
FHPIBS	0,78	0,71	0,66	0,64	0,62	0,60	0,58
FHPIBD	0,78	0,72	0,71	0,66	0,64	0,61	0,59

en correspondance Dense).

La deuxième version utilise la méthode d'estimation mono-couche dérivée des estimateurs en situation de bi-transparence, qui a déjà été présentée en 5.5.1.2. Nous choisissons de ne pas effectuer la troisième étape facultative d'estimation paramétrique, car nous avons vu en 5.5.1.2 qu'elle peut nuire à la précision de l'estimation dans ce cas. Le filtre en question est dénommé *FTACMP* (Filtre Temporel Adaptatif Compensé en mouvement Mono-couche, estimés par mise en correspondance par blocs, puis extraction des Paramètres de mouvement).

- le **filtre temporel hybride** (par patch itéré FHPI) qui s'appuie sur les résultats de l'estimation des mouvements transparents en situation de bi-transparence simple (chapitre 3 - FHPIBS⁷), ou bien en situation de bi-transparence distribuée (chapitre 5 - FHPIBD⁸).

L'évolution des écarts-type du bruit résiduel de la séquence restaurée pour ces quatre filtres est donnée dans le tableau 8.4. Nous y remarquons que les deux types de filtres se comportent de la même façon pour les deux estimateurs de mouvement utilisés. Le filtre adaptatif compensé en mouvement monocouche atteint en régime stationnaire un taux de débruitage de 35% (soit un PSNR de 25.9dB). Le filtre hybride est plus efficace et va jusqu'à un taux de débruitage de 42% (soit un PSNR de 26,8dB).

Nous présentons dans le tableau 8.5 l'écart entre la vérité terrain et la moyenne de 25 images débruitées respectivement pour les méthodes FTACMP et FHPIBD. Comme expliqué plus haut, ceci permet de quantifier la préservation de l'information utile puisque le bruit est fortement atténué par l'opération de moyennage [Aach 00].

Nous y observons une plus grande fidélité aux données que lorsque les mouvements sont parfaitement connus : la métrique vaut 4.37 lorsque les mouvements ont été estimés, contre 5.36 lorsqu'ils sont connus. Cette constatation en apparence paradoxale s'explique par le fait que nos données sont simulées, ce qui implique quelques légers biais dus notamment

⁷FHPI, avec mouvements estimés en situation de Bitransparence Simple.

⁸FHPI, avec mouvements estimés en situation de Bitransparence Distribuée.

TAB. 8.5: Écart quadratique moyen entre la vérité terrain et la moyenne de 25 images débruitées pour différentes méthodes. Ceci permet de quantifier la qualité de la conservation de l'information pertinente. Les images considérées sont entachées d'un bruit typique d'images fluoroscopiques, d'écart-type $\sigma = 20$.

FTACMP	FHPIBD
4.96	4.37

aux problèmes d'interpolations⁹. Une estimation correcte des mouvements dans ce cadre pourra aboutir à des mouvements légèrement plus pertinents au sens de l'ECMT qu'une utilisation directe de la vérité-terrain.

D'autre part, nous remarquons un meilleur comportement du FHPIBD que du FTACMP vis à vis de cette métrique : un PSNR de 35.32 dB contre 34.22 dB. Ce résultat est logique car l'estimation des mouvements du FTACMP ne modélise pas la transparence, et ne peut donc mener à des mouvements respectant aussi bien l'ECMT que ceux qui sont issus de l'estimation du chapitre 5.

Nous présentons également à la figure 8.8 la huitième image de la séquence débruitée dans les deux types d'approches (mouvements mono-couches ou transparents). Les images résultantes sont visuellement équivalentes, et nous remarquons notamment que le filtre adaptatif mono-couche conserve aussi bien¹⁰ les contrastes des objets très contrastés que le filtre temporel hybride^{11 12}. La différence avec le filtre adaptatif non compensé est par contre éloquente.

8.6 Résultats sur séquences réelles

Il est délicat d'obtenir et de mettre en valeur des résultats éloquentes du débruitage proposé sur des séquences cliniques réelles pour deux raisons :

- la qualité du filtrage dépend étroitement des mouvements transparents estimés (selon la méthode proposée dans le chapitre 5). Or, cette estimation est forcément imparfaite, ce qui est problématique étant donné la structure récursive du filtre : le filtre hybride n'atteindra sa pleine capacité de débruitage que si les mouvements anatomiques sont bien estimés sur plusieurs instants consécutifs.

D'autre part, les estimations de mouvements transparents s'appuyant sur l'ECMT supposent les mouvements constants sur deux intervalles de temps consécutifs, ce qui est faux à certains instants du cycle cardiaque. Les estimés de mouvements obtenus sont donc nécessairement erronés à ces dates, ce qui limite la capacité de débruitage du filtre.

⁹D'autant plus que les images originales correspondant aux couches sont elles aussi bruitées, et donc très sensibles aux problèmes d'interpolation.

¹⁰Et même légèrement mieux.

¹¹Sur les objets très contrastés, l'information provenant de leur couche domine tellement celle de l'autre couche qu'il est possible d'appréhender ses mouvements avec précision sans modéliser la transparence.

¹²Le filtre hybride permet toutefois, outre cette conservation des forts contrastes, une réduction de bruit plus importante, et préserve mieux les objets faiblement contrastés.

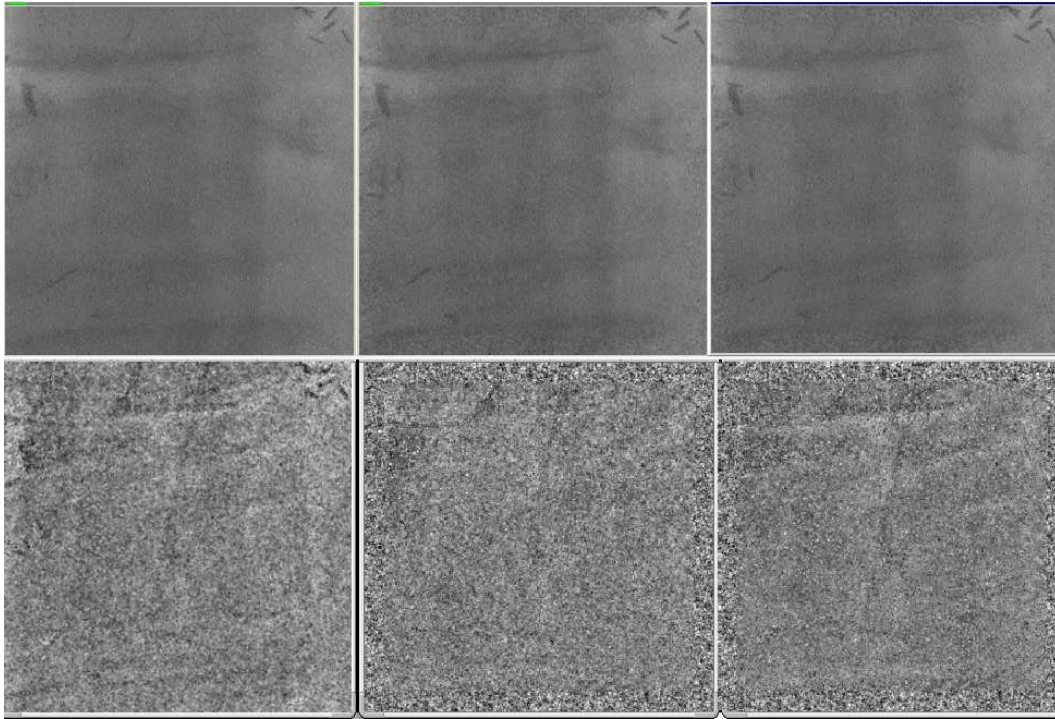


FIG. 8.8: En haut, la huitième image de la séquence, en bas, sa différence avec l'image non bruitée correspondante. De gauche à droite : le filtre adaptatif non compensé en mouvement (référence), le filtre adaptatif compensé en mouvement monocouche (estimé avec la version paramétrique) et le filtre hybride (estimé avec la méthode en situation de bitransparence distribuée). Les bords de l'image ne sont pas filtrés.

- il est difficile de rendre compte d'un gain de débruitage ou d'une amélioration de conservation des contrastes sur des images statiques imprimées. Les séquences d'images fluoroscopiques sont observées par le médecin sur des écrans spécifiques dans une salle plongée dans la pénombre. Il est certain qu'un observateur est bien plus sensible au bruit et au contraste des objets d'intérêt dans de telles conditions que lorsqu'il doit se prononcer sur des images statiques imprimées.

Pour ces deux raisons, nous avons extrait les images les plus démonstratives des tests que nous avons effectués afin de valider notre approche. Nous avons choisi de présenter en vis à vis les résultats du filtrage hybride (dans sa version FHPI, à gauche des figures présentées) et ceux du filtrage temporel récursif sans compensation de mouvement FTA (à droite). Ces deux filtres ont été réglés pour avoir un niveau de débruitage équivalent, de sorte que sur c'est sur leur capacité à préserver les détails et à n'introduire que peu d'artefacts que nous allons les juger.

La figure 8.9 se concentre sur un détail de la séquence 3, déjà présentée à la figure 3.18. La masse sombre du cœur (à droite) se détache clairement des tissus clairs des poumons (à gauche), et le tout se superpose à un fond statique matérialisé par les disques de la colonne vertébrale.

Nous présentons quatre images de cette séquence, traitée par le FHPI et le FTA. La comparaison des images en vis à vis met en évidence une bien meilleure conservation des contrastes de la masse du cœur. Cet effet est confirmé par l'observation des tissus des poumons.

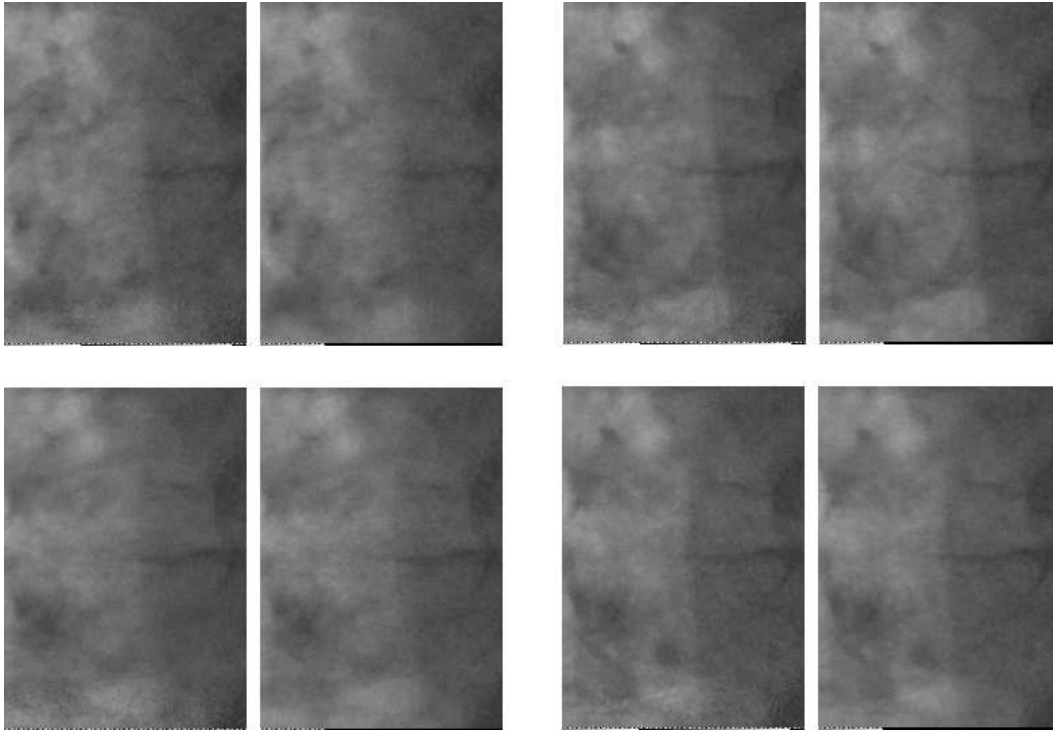


FIG. 8.9: Quatre instants de la séquence d'images fluoroscopiques 3, traitées par le FHPI (à gauche des sous-images) et le FTA (à droite). Nous observons une meilleure conservation des contrastes par le filtre hybride, en particulier au niveau de la frontière du cœur.

Les deux figures de résultats suivants (Fig.8.10 et 8.11) ont été obtenues sur la séquence fluoroscopique 6, déjà utilisée à la figure 5.7. Il s'agissait d'un examen cardiaque sur lequel la méthode d'estimation de mouvements transparents du chapitre 5 parvenait à appréhender les mouvements du cathéter. Nous observons en effet sur la figure 8.10 que le contraste du cathéter est plus marqué sur les images traitées par le FHPI que par le FTA pour les quatre instants présentés.

D'autre part, nous nous arrêtons à la Fig.8.11 sur l'artefact de "queue de bruit" produit par le FTA. Quand ce filtre détecte le mouvement d'un objet texturé, il réintroduit une partie de l'image originale, bruitée, au milieu de l'image débruitée. Dans ces mêmes conditions, le filtre hybride sera en mesure de continuer à filtrer, et donc n'introduit pas ce type d'artefact (pour peu que les mouvements soient bien estimés).

Nous observons ce phénomène à la figure 8.11. Là encore, il est moins visible sur une image statique imprimée que sur une séquence dynamique d'images observées dans de bonnes conditions. Nous présentons donc à côté de la séquence traitée un zoom sur la partie d'intérêt, au centre de l'image. Nous y observons bien que l'arc correspondant à la frontière du

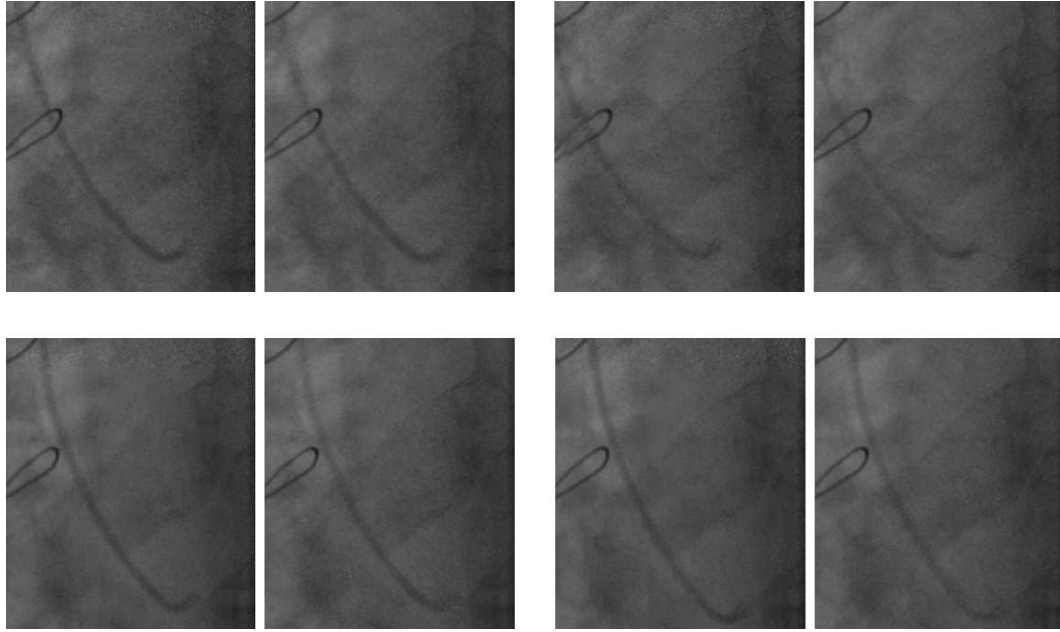


FIG. 8.10: Quatre instants de la séquence d'images fluoroscopiques 6, traitées par le FHPI (à gauche des sous-images) et le FTA (à droite). Nous observons une meilleure conservation des contrastes du cathéter par le filtre hybride.

cœur en mouvement (diagonale sombre au centre de l'image dans le sens de la longueur) est bruité dans l'image correspondant au FTA. Un artefact de ce type n'est pas présent dans l'image traitée par le FHPI.

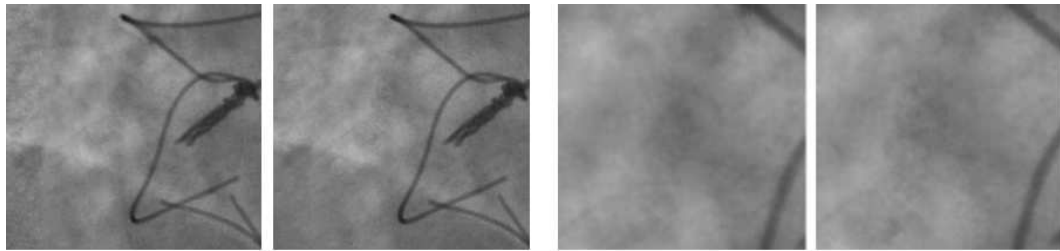


FIG. 8.11: Image de la séquence d'images fluoroscopiques 6, traitée par le FHPI (à gauche) et le FTA (à droite). Les deux images de droite correspondent à un zoom mettant en évidence l'artefact de "queue de bruit" produit par le FTA.

Enfin, les résultats présentés à la figure 8.12 concernent la séquence fluoroscopique 1, que nous avons introduite à la figure 3.20. Nous y observons des agrafes chirurgicales mues par le mouvement de la respiration. Elles se superposent au diaphragme. Les résultats correspondant aux trois premiers instants présentés mettent en évidence une bien meilleure conservation du contraste des agrafes par le FHPI que par le FTA. Cependant, nous montrons sur les résultats correspondant au quatrième instant présenté que ces

conclusions peuvent être inversées lorsque les estimations des mouvements sont erronées.

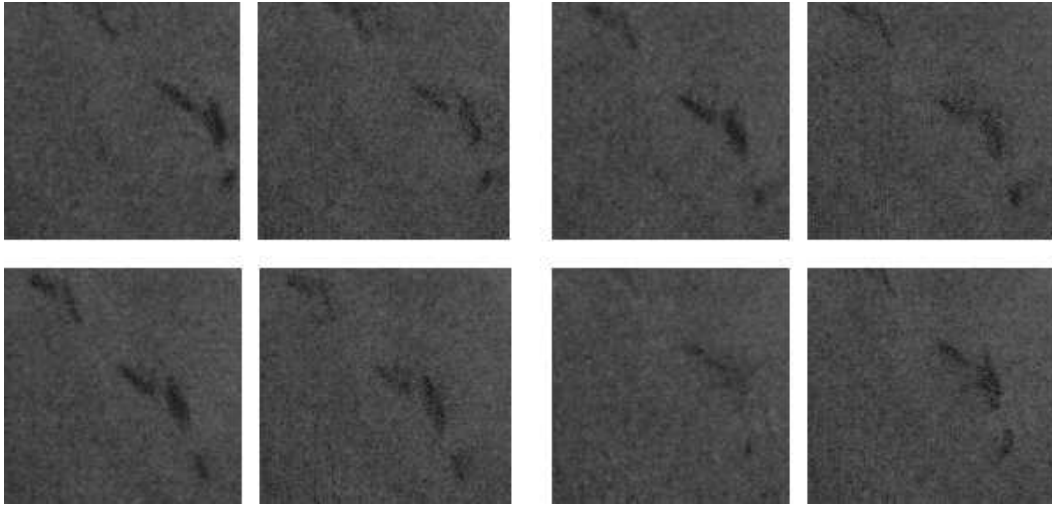


FIG. 8.12: Quatre instants de la séquence d’images fluoroscopiques 1, traitées par le FHPI (à gauche des sous-images) et le FTA (à droite). Nous observons une meilleure conservation des contrastes des agrafes par le filtre hybride lorsque les mouvements estimés sont pertinents (trois premiers instants présentés). Par contre, ces conclusions peuvent être inversées quand les estimés des mouvements sont erronés.

8.7 Conclusion

Nous avons présenté dans ce chapitre un filtre temporel dit “hybride” qui reprend le principe du filtre compensé en mouvements transparents quand les deux couches sont texturées, mais se contente de suivre la couche la plus informative quand une des deux couches est homogène. Ce faisant, il est possible de dépasser la limite de débruitage du filtre compensé en mouvements transparents sans perdre d’information.

La décision sur l’homogénéité des régions est similaire au principe du calcul des gains des filtres adaptatifs, de sorte que l’image filtrée finale ne s’appuie pas sur des décisions binaires mais sur une pondération des images filtrées pour les différentes configurations considérées. La méthode inclut l’étude de la qualité des estimés des mouvements, ce qui est nécessaire pour garder des résultats acceptables quand les mouvements ne sont pas connus.

Ce filtre permet une excellente conservation des contrastes et une réduction de bruit nettement supérieure aux autres filtres temporels sur des images en mouvement. Différents résultats sur des séquences cliniques valident son apport : il permet un meilleur compromis débruitage-conservation de l’information qu’un filtre temporel non compensé en mouvement. On peut d’ailleurs jouer sur l’équilibre entre ces deux propriétés en filtrant de manière plus ou moins marquée. Notons en outre qu’un filtre adaptatif compensé en mouvement non transparent peut constituer une première étape intéressante de filtrage compensé en mouvement au niveau industriel.

La limitation majeure de ce type de stratégie est la dépendance des filtres proposés à la

qualité de l'estimation des mouvements. En particulier, les performances de débruitage sont notablement dégradées par les instants du cycle cardiaque pour lesquelles les hypothèses d'application de l'ECMT ne sont pas valides. Nous y reviendrons dans les perspectives de cette thèse.

Si les performances théoriques de ce filtre dépassent celles des autres filtres temporels, il ne permet quand même pas d'approcher les images idéales, même lorsque les vrais mouvements sont connus. C'est pourquoi nous allons relâcher la contrainte concernant la conservation de l'esthétique de l'image pour étudier un filtrage *spatio-temporel* avec compensation des mouvements transparents dans le chapitre suivant.

Chapitre 9

Débruitage spatio-temporel compensé selon les mouvements transparents

Nous étendons dans la première section de ce chapitre la compensation des mouvements transparents au cas des filtres spatio-temporels. Les performances de cette classe de filtres étant liées à la stationarité temporelle de la séquence, ils sont aussi susceptibles de tirer profit d'une compensation des mouvements transparents. Même si l'utilisation de filtres spatio-temporels se heurte au problème "d'esthétique" du bruit résiduel, leur étude est intéressante car ils permettent d'atteindre des niveaux de débruitage supérieurs aux filtres temporels.

Nous montrons en étudiant le cas particulier d'un filtre spatio-temporel adaptatif que, comme dans le cas des filtres *temporels*, le niveau de débruitage permis par le filtre compensé en mouvement est contraint par le fait que la compensation de mouvements transparents augmente le bruit des images recalées.

Nous proposons alors dans la deuxième section d'adapter le filtrage hybride aux filtres spatio-temporels.

9.1 Filtres spatio-temporels compensés en mouvement

9.1.1 Principe

Les filtres spatio-temporels sont d'autant plus efficaces qu'ils sont utilisés sur des supports stationnaires. La compensation des mouvements transparents est donc susceptible d'améliorer leur performance.

Nous pouvons construire des filtres spatio-temporels compensés en mouvements transparents de la même façon que dans le chapitre 7, ce que nous avons schématisé dans le tableau 9.1. Nous substituons aux images $\hat{I}(\mathbf{p}, t - n), \dots, \hat{I}(\mathbf{p}, t - 1)$, normalement utilisées par le filtre¹, les images correspondantes compensées en mouvements transparents $\bar{I}_t(\mathbf{p}, t - n)$,

¹Avec bien sûr l'image acquise à la date t , $I(\mathbf{p}, t)$.

..., $\bar{I}_t(\mathbf{p}, t - 1)$, selon :

$$\begin{aligned} \bar{I}_t(\mathbf{p}, t - k) &= \hat{I}(\mathbf{p} + k\mathbf{w}_1, t - k) + \hat{I}(\mathbf{p} + k\mathbf{w}_2, t - k) \\ &\quad - \hat{I}(\mathbf{p} + k\mathbf{w}_1, +k\mathbf{w}_2, t - 2k) \quad \text{pour } k \in [1, n]. \end{aligned} \quad (9.1)$$

Nous reconnaissons en (9.1) l'équation de contrainte du mouvement transparent appliquée entre les instants $t - 2k$, $t - k$ et t . Elle suppose que les mouvements des couches sont constantes² entre $t - 2k$ et t . Cette hypothèse, qui devient très forte quand le support temporel du filtre est étendu, n'est autre que la condition "habituelle" de constance des mouvements sur deux intervalles de temps consécutifs utilisée en bitransparence quand le filtre n'utilise qu'une image passée.

Dans les hypothèses d'application de (9.1), la stationnarité temporelle est vérifiée sur $\bar{I}_t(\mathbf{p}, t - n)$, ..., $\bar{I}_t(\mathbf{p}, t - 1)$ et $I(\mathbf{p}, t)$. Cependant, nous avons vu dans le chapitre 7 que les images compensées $\bar{I}_t(\mathbf{p}, t - k)$ ont par ailleurs l'inconvénient d'être plus bruitées que les images originelles. Les deux approches possibles sont donc les suivantes :

- **filtres spatio-temporels sans compensation des mouvements transparents ;**
les images considérées sont moins bruitées *mais* non stationnaires temporellement.
- **filtres spatio-temporels avec compensation des mouvements transparents ;**
les images considérées sont stationnaires temporellement *mais* bruitées.

La première approche conserve mieux les détails (car aucun lissage temporel ne pourra les atténuer) mais diminue moins le bruit que la seconde.

TAB. 9.1: Compensation des mouvements transparents pour les filtres spatio-temporels

Filtre	Spatio-temporel simple	Spatio-temporel avec compensation des mouvements
$\tilde{I}(\mathbf{p}, t)$	$f(\tilde{I}(\mathbf{p}, t - n), \dots, \tilde{I}(\mathbf{p}, t - 1), I(\mathbf{p}, t))$	$f(\bar{I}_t(\mathbf{p}, t - n), \dots, \bar{I}_t(\mathbf{p}, t - 1), I(\mathbf{p}, t))$
Avantage	S'appuie sur des images moins bruitées	S'appuie sur des images stationnaires temporellement

9.1.2 Résultats sur séquence synthétique

Filtre bilatéral étendu. Nous étudions le comportement d'un cas particulier de filtre spatio-temporel, auquel nous adjoignons une compensation des mouvements transparents. Nous avons choisi une variation du filtre bilatéral [Tomasi 98]. Nous l'avons étendu au

²Dans ces conditions, la couche 1 s'est déplacée entre $t - k$ et t de $k\mathbf{w}_1$ et la couche 2 de $k\mathbf{w}_2$.

traitement de séquences d'images comme suit. L'image filtrée est donnée par :

$$\hat{I}(\mathbf{p}, t) = \frac{1}{W} \left(\sum_{\mathbf{p}' \in V(\mathbf{p})} w_1(\mathbf{p}, \mathbf{p}') I(\mathbf{p}', t) + \sum_{\mathbf{p}' \in V(\mathbf{p})} w_2(\mathbf{p}, \mathbf{p}') \hat{I}(\mathbf{p}', t-1) \right) \quad (9.2)$$

$$\text{avec } W = \sum_{\mathbf{p}' \in V(\mathbf{p})} w_1(\mathbf{p}, \mathbf{p}') + \sum_{\mathbf{p}' \in V(\mathbf{p})} w_2(\mathbf{p}, \mathbf{p}') \quad (9.3)$$

$$w_1(\mathbf{p}, \mathbf{p}') = \exp \left(- \frac{(I(\mathbf{p}, t) - I(\mathbf{p}', t))^2}{(2Z\sigma(\mathbf{p}, t))^2} \right) \quad (9.4)$$

$$w_2(\mathbf{p}, \mathbf{p}') = \exp \left(- \frac{(I(\mathbf{p}, t) - \hat{I}(\mathbf{p}', t-1))^2}{Z^2(\sigma^2(\mathbf{p}, t) + \sigma_f^2(\mathbf{p}, t-1))} \right), \quad (9.5)$$

où $V(\mathbf{p})$ désigne un voisinage spatial de \mathbf{p} (en pratique 3×3 ou 5×5) et Z est un facteur qui influe sur l'adaptation des poids (en pratique, $Z = 2$ pour un filtrage conservatif, $Z = 3$ pour un filtrage plus marqué). Remarquons que l'argument du terme exponentiel défini en (9.4) et (9.5) est normé par l'écart-type du bruit entachant $I(\mathbf{p}, t) - I(\mathbf{p}', t)$ et $I(\mathbf{p}, t) - \hat{I}(\mathbf{p}', t-1)$.

Il s'agit donc d'un filtre spatio-temporel adaptatif qui moyenne les pixels d'un voisinage spatio-temporel, en les pondérant par un facteur qui sera élevé si les deux niveaux de gris considérés sont proches (et font donc probablement partie d'une même région stationnaire), et faible s'ils sont nettement différents.

Filtre bitatéral étendu avec compensation des mouvements transparents. La version de ce filtre avec compensation des mouvements transparents se déduit de la formule (9.2) en remplaçant les images passées débruitées $\hat{I}(\cdot, t-1)$ par les images prédites $\bar{I}(\cdot, t)$:

$$\hat{I}(\mathbf{p}, t) = \frac{1}{W} \left(\sum_{\mathbf{p}' \in V(\mathbf{p})} w_1(\mathbf{p}, \mathbf{p}') I(\mathbf{p}', t) + \sum_{\mathbf{p}' \in V(\mathbf{p})} w_2(\mathbf{p}, \mathbf{p}') \bar{I}_t(\mathbf{p}', t-1) \right) \quad (9.6)$$

$$\text{avec } W = \sum_{\mathbf{p}' \in V(\mathbf{p})} w_1(\mathbf{p}, \mathbf{p}', t-1) + \sum_{\mathbf{p}' \in V(\mathbf{p})} w_2(\mathbf{p}, \mathbf{p}', t) \quad (9.7)$$

$$w_1(\mathbf{p}, \mathbf{p}') = \exp \left(- \frac{(I(\mathbf{p}, t) - I(\mathbf{p}', t))^2}{(2Z\sigma(\mathbf{p}, t))^2} \right) \quad (9.8)$$

$$w_2(\mathbf{p}, \mathbf{p}') = \exp \left(- \frac{(I(\mathbf{p}, t) - \bar{I}_t(\mathbf{p}', t-1))^2}{Z^2(\sigma^2(\mathbf{p}, t) + \bar{\sigma}_t^2(\mathbf{p}, t-1))} \right). \quad (9.9)$$

Résultats expérimentaux. Nous présentons dans le tableau 9.3 les niveaux des bruits résiduels après traitement par le filtre bilatéral étendu, compensé en mouvement (CM) ou non. La séquence synthétique utilisée est la même que celle du chapitre précédent (deux couches cardiaques, une en translation, l'autre en déplacement homothétique, simulant un examen fluoroscopique). Nous testons les réglages $Z = 2$ et $Z = 3$, et les voisinages 3×3 et 5×5 .

Nous pouvons remarquer que les niveaux de débruitage atteints par ce filtre sont plus élevés que dans le cas du filtre temporel adaptatif : de l'ordre de 50% de réduction de l'écart-type pour un voisinage 3×3 (PSNR 28.1dB) et de 60% pour un voisinage 5×5 (PSNR 30.1dB). D'autre part, le filtre réduit le bruit dès la première image (numérotée 0) de la

TAB. 9.2: Evolution du niveau de bruit résiduel normalisé $\sigma_f(\mathbf{p}, t)/\sigma$ d'une séquence réaliste débruitée par différentes versions du filtre bilatéral étendu : avec compensation des mouvements transparents (CM) ou sans, $Z = 2$ ou 3 , voisinages 3×3 ou 5×5 (le mouvement des deux couches est connu).

CM	Z	vois	0	1	2	3	4	5	6	7	8
non	2	3	0,61	0,49	0,49	0,49	0,48	0,49	0,49	0,49	0,49
non	3	3	0,58	0,46	0,46	0,46	0,47	0,48	0,48	0,48	0,48
non	2	5	0,47	0,40	0,40	0,41	0,41	0,42	0,42	0,42	0,42
non	3	5	0,44	0,39	0,40	0,41	0,42	0,43	0,44	0,44	0,45
oui	2	3	0,61	0,50	0,61	0,63	0,62	0,63	0,63	0,63	0,63
oui	3	3	0,58	0,46	0,56	0,58	0,57	0,58	0,58	0,58	0,58
oui	2	5	0,47	0,40	0,49	0,48	0,49	0,49	0,49	0,49	0,49
oui	3	5	0,44	0,39	0,43	0,44	0,44	0,44	0,45	0,44	0,45

séquence du fait de sa composante spatiale. Par ailleurs, la compensation des mouvements transparents ne devient effective qu'à partir de la troisième image de la séquence

Par la suite, le filtre réduit moins efficacement le bruit que sa version non compensée, ce qui est tout à fait logique étant donné qu'il exploite les images $\bar{I}_t(\mathbf{p}, t - k)$ plus bruitées que les $\hat{I}(\mathbf{p}, t - k)$.

La figure 9.1 affichant les bruits résiduels met en évidence les artefacts liés à ce type de filtrage. La corrélation spatiale du bruit résiduel y est évidente, notamment sur les images résultant de l'application du filtre bilatéral avec des voisinages 5×5 . D'autre part, il est clair que la méthode sans compensation de mouvement induit des pertes notables de contrastes (surtout dans le cas du voisinage 5×5), alors que l'information pertinente est conservée lorsque l'on compense les mouvements transparents.

Des huit traitements menés, nous pouvons conclure que

- c'est le filtre bilatéral étendu compensé en mouvement, appliqué sur un voisinage 5×5 et pour $Z = 3$, qui donne les meilleurs résultats si on fait abstraction de l'esthétique du bruit. L'écart-type du bruit est réduit de 55%.
- c'est le filtre bilatéral étendu compensé en mouvement, appliqué sur un voisinage 3×3 et pour $Z = 3$, qui donne les meilleurs résultats si on prend en compte l'esthétique du bruit. Un observateur restera cependant probablement gêné par la granularité du bruit et préférera un filtrage temporel pur. L'écart-type du bruit est réduit de 42%.

Ces expériences confirment donc les conclusions énoncées plus haut : **le filtre compensé en mouvements transparents conserve les contrastes, mais débruite moins fortement les images que la version sans compensation de mouvements transparents.**

9.2 Filtres spatio-temporels hybrides

9.2.1 Principe

Le filtre hybride permet de proposer un autre compromis "images stationnaires contre images peu bruitées". L'idée est la même que dans le chapitre 8 : puisque les images cli-

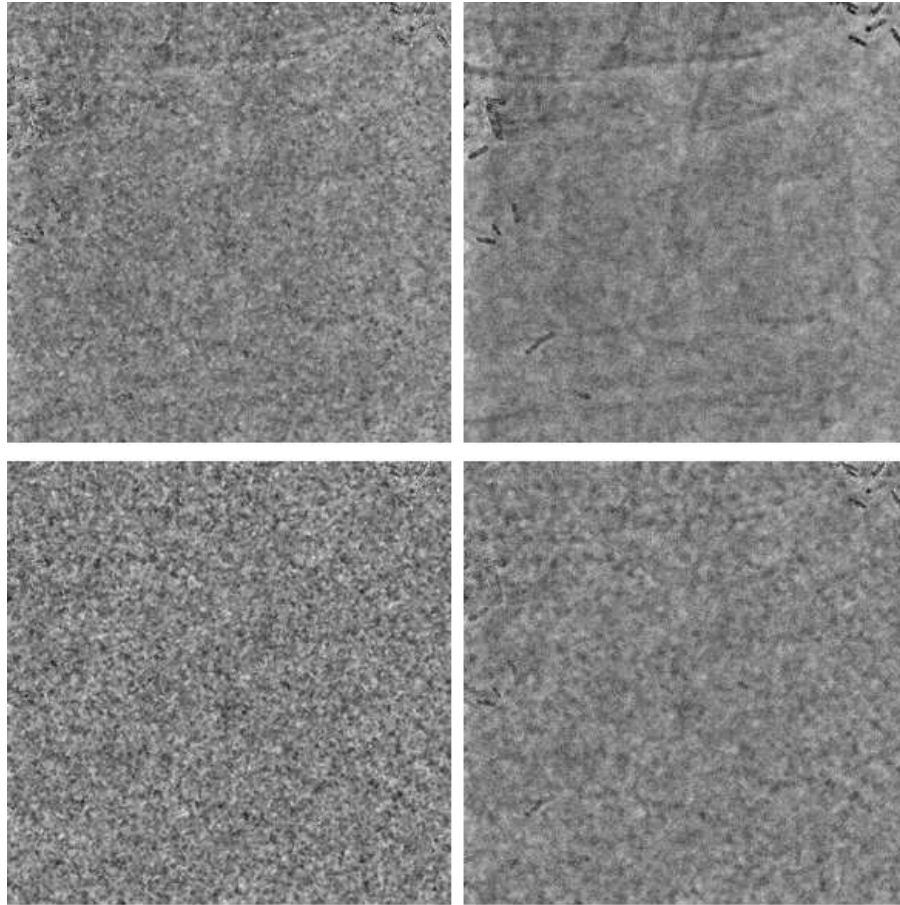


FIG. 9.1: Carte des bruits résiduel correspondant à la huitième image de la séquence débruitée. *En haut* : application du filtre bilatéral étendu sans compensation des mouvements transparents, à gauche avec le réglage $Z = 2$ pour un voisinage 3×3 , et à droite avec le réglage $Z = 3$ pour un voisinage 5×5 . *En bas* : application du filtre bilatéral étendu avec compensation des mouvements, à gauche avec le réglage $Z = 2$ pour un voisinage 3×3 , et à droite avec le réglage $Z = 3$ pour un voisinage 5×5 .

niques que nous considérons sont homogènes sur de larges régions, la compensation de mouvements transparents peut être réservée aux pixels correspondant à deux couches hétérogènes.

Comme dans le chapitre précédent, nous ne prenons pas de décision définitive sur la configuration observée en chaque pixel. Nous appliquons plutôt le filtre spatio-temporel dans les quatre configurations locales possibles :

- (\mathbf{C}_0) : les deux couches sont hétérogènes. Les images aux dates précédentes, entrant en jeu dans le filtre sont compensées selon les mouvements transparents estimés ;
- (\mathbf{C}_1) : seule la première couche est hétérogène. Les images précédentes sont compensées selon le mouvement de cette couche ;
- (\mathbf{C}_2) : seule la deuxième couche est hétérogène. Les images précédentes sont com-

pensées selon le mouvement de cette couche ;

- (C₃) : les deux couches sont considérés comme homogènes. Nous choisissons de compenser le mouvement de la couche la plus texturée³.

Nous ajoutons l'étude d'une configuration portant sur la qualité de l'estimation des mouvements :

- (C₄) : les mouvements estimés sont erronés. Nous utilisons alors $I(\mathbf{p}, t)$ et lui appliquons un filtrage spatial.

Les cinq images résultantes $\hat{I}_{(C_0)}$, $\hat{I}_{(C_1)}$, $\hat{I}_{(C_2)}$, $\hat{I}_{(C_3)}$ et $\hat{I}_{(C_4)}$ sont alors combinées selon une approche voisine des filtres adaptatifs, en utilisant la formule (8.13) que nous rappelons ici :

$$\begin{aligned}
 \hat{I}(\mathbf{p}, t) = & h_{12}(\mathbf{p})(1 - h_1(\mathbf{p}))(1 - h_2(\mathbf{p}))\hat{I}_{(C_0)}(\mathbf{p}, t) \\
 & + h_{12}(\mathbf{p})(1 - h_1(\mathbf{p}))h_2(\mathbf{p})\hat{I}_{(C_1)}(\mathbf{p}, t) \\
 & + h_{12}(\mathbf{p})h_1(\mathbf{p})(1 - h_2(\mathbf{p}))\hat{I}_{(C_2)}(\mathbf{p}, t) \\
 & + h_{12}(\mathbf{p})h_1(\mathbf{p})h_2(\mathbf{p})\hat{I}_{(C_3)}(\mathbf{p}, t) \\
 & + (1 - h_{12}(\mathbf{p}))\hat{I}_{(C_4)}(\mathbf{p}, t
 \end{aligned} \tag{9.10}$$

Les facteurs h_1 , h_2 et h_{12} sont spécifiés comme dans le chapitre précédent par seuillage doux de $|I(\mathbf{p}, t) - \hat{I}(\mathbf{p} + \mathbf{u}, t)|$, de $|I(\mathbf{p}, t) - \hat{I}(\mathbf{p} + \mathbf{v}, t)|$ et de $|I(\mathbf{p}, t) - \tilde{I}(\mathbf{p}, t)|$. Puisque les filtres spatio-temporels produisent des débruitages relativement marqués, nous n'envisageons en général ni prétraitement par application du filtre bilatéral, ni par patch.

L'implantation pratique de ce filtre repose sur un modèle d'évolution du bruit résiduel le long de la séquence pour chaque filtre spatio-temporel dont nous dérivons une version hybride⁴. Nous résumons la méthode proposée à la figure 9.2.

9.2.2 Résultats sur séquences synthétiques

Nous étudions le comportement des filtres spatio-temporels hybrides avec l'utilisation du filtre spatial bilatéral étendu présenté dans la section 9.1.2. Nous avons traité la séquence synthétique utilisée en section 9.1.2 par un filtre bilatéral étendu hybride pour différents réglages. Nous présentons dans le tableau 9.3 les niveaux de bruit résiduel des images traitées par ce filtre (*FSTH* Filtre Spatio-Temporel Hybride), et par le filtre bilatéral étendu avec compensation des mouvements transparents de la section précédente (*FSTCM* Filtre Spatio-Temporel Compensé en Mouvements Transparentes).

Nous notons que le FSTH permet un débruitage plus marqué que la version avec compensation des mouvements transparents (FSTCM) : entre 5% ($\Delta PSNR = 0.7dB$ pour le cas du voisinage 3×3) et 3% de bruit ($\Delta PSNR = 0.6dB$) supplémentaire est filtré. Les performances de la version hybride FSTH en terme de réduction de l'écart-type du bruit sont par contre plus faibles que celle du filtre non compensé en mouvement (FST Filtre Spatio-Temporel direct).

³Ce que l'on évalue en comparant $|I(\mathbf{p}, t) - \hat{I}(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t - 1)|$ et $|I(\mathbf{p}, t) - \hat{I}(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t - 1)|$. À la différence du chapitre précédent, nous préférons ne pas nous appuyer sur l'ensemble des images qui seraient pertinentes ici ($\hat{I}(\mathbf{p} + \mathbf{w}_1(\mathbf{p}), t - 1)$, $\hat{I}(\mathbf{p} + \mathbf{w}_2(\mathbf{p}), t - 1)$ et $\tilde{I}(\mathbf{p}, t)$) pour éviter un filtrage trop marqué qui risquerait de faire disparaître certaines information.

⁴Pour ce faire, nous nous sommes basés sur des exemples synthétiques représentatifs des séquences à traiter.

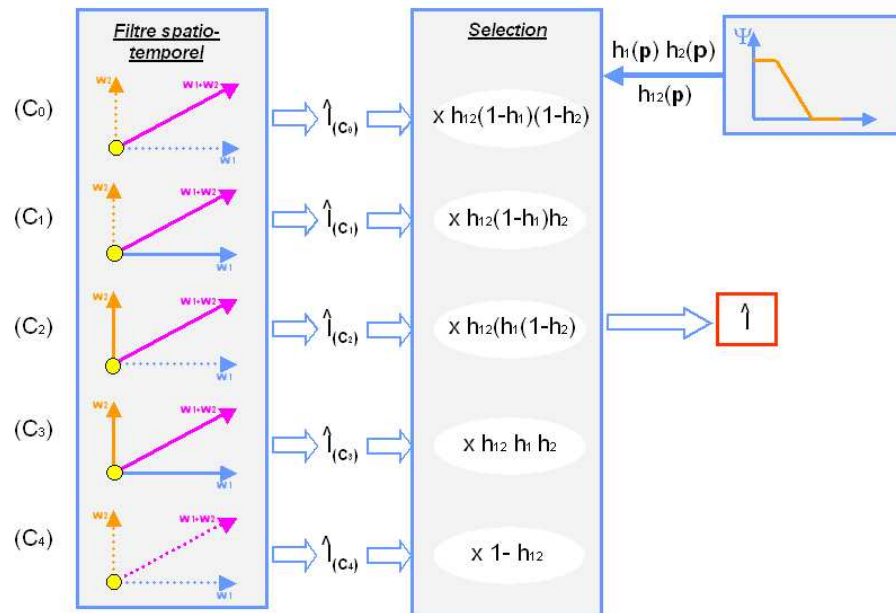


FIG. 9.2: Schéma de fonctionnement du filtre spatio-temporel hybride. Il est composé de deux modules : le premier filtre de manière spatio-temporelle les images dans les cinq configurations possibles, le second pondère les cinq images intermédiaires selon la “probabilité” de leur occurrence au pixel considéré.

Pour juger de la qualité de la conservation de l’information, nous présentons dans le tableau 9.4 l’écart quadratique moyen entre la vérité terrain d’une part, et la moyenne de 25 images débruitées pour ces différentes méthodes d’autre part. Ce faisant, l’impact du bruit est fortement réduit et l’on peut juger quantitativement de la conservation globale des contrastes [Aach 00].

Nous y observons une plus grande distorsion à la vérité terrain quand le filtre bilatéral étendu est réglé de manière agressive (voisinage 3×3 , $Z = 2$) que lorsqu’il filtre de manière moins marquée (vois 5×5 , $Z = 3$), et ce quelle que soit la stratégie de compensation. D’autre part, le FSTH conserve bien mieux l’information efficace que le FST. Cependant c’est bien évidemment le FSTCM qui se révèle le plus fidèle aux images idéales, au prix d’un débruitage moins marqué.

Les cartes des bruits résiduels sont affichées à la figure 9.3. Elles correspondent à la huitième image pour des applications des filtres FST, FSTCM et FSTH, pour $Z = 2$ avec un voisinage 3×3 , et $Z = 3$ avec un voisinage 5×5 .

Nous remarquons une bien meilleure conservation des contrastes dans le cas FSTH que dans le cas FST, même si elle n’est pas aussi bonne que celle permise par le FSTCM⁵. Par contre, l’artefact de coloration du bruit est évidemment également perceptible dans les trois configurations.

⁵Ce qui est normal car ce dernier garantit une conservation parfaite des contrastes.

TAB. 9.3: Evolution du niveau de bruit résiduel normalisé $\sigma_f(\mathbf{p}, t)/\sigma$ d'une séquence réaliste débruitée pour différentes versions du filtre bilatéral étendu : avec compensation des mouvements transparents (CM) ou sans, $Z = 2$ ou 3 , voisinages 3×3 ou 5×5 (le mouvement des deux couches est connu).

	Z	vois	0	1	2	3	4	5	6	7	8
CM	2	3	0,61	0,50	0,61	0,63	0,62	0,63	0,63	0,63	0,63
CM	3	3	0,58	0,46	0,56	0,58	0,57	0,58	0,58	0,58	0,58
CM	2	5	0,47	0,40	0,49	0,48	0,49	0,49	0,49	0,49	0,49
CM	3	5	0,44	0,39	0,43	0,44	0,44	0,44	0,45	0,44	0,45
H	2	3	0,61	0,50	0,57	0,58	0,57	0,58	0,58	0,58	0,58
H	3	3	0,58	0,41	0,52	0,53	0,52	0,53	0,53	0,54	0,53
H	2	5	0,47	0,42	0,46	0,46	0,46	0,46	0,46	0,46	0,46
H	3	5	0,44	0,41	0,42	0,42	0,42	0,43	0,43	0,43	0,42

TAB. 9.4: Écart quadratique moyen entre la vérité terrain et la moyenne de 25 images débruitées pour différentes méthodes. Ceci permet de quantifier la qualité de la conservation de l'information pertinente. Les images considérées sont entachées d'un bruit typique d'images fluoroscopiques, d'écart-type $\sigma = 20$.

	FST	FSTCM	FSTH
vois 3×3 , $Z = 2$	4.97	3.63	4.24
vois 5×5 , $Z = 3$	9.71	4.33	5.86

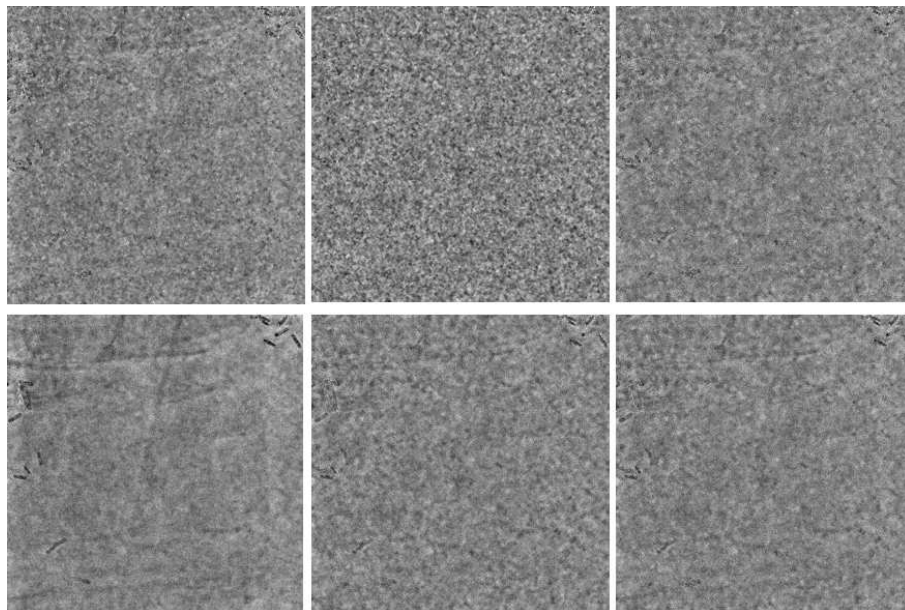


FIG. 9.3: Cartes des bruits résiduels correspondant à la huitième image de la séquence débruitée. *En haut* : filtre bilatéral avec $Z = 2$ sur un voisinage 3×3 , à gauche sans compensation de mouvement (FST), au milieu avec compensation de mouvement (FSTCM) et à droite dans une version hybride (FSTH). *En bas* : filtre bilatéral $Z = 3$ avec un voisinage 5×5 , à gauche sans compensation de mouvement (FST), au milieu avec compensation de mouvement (FSTCM) et à droite dans une version hybride (FSTH).

9.3 Conclusion

Nous avons étendu dans ce chapitre la stratégie des compensation de mouvements transparents et de filtrage hybride aux filtres spatio-temporels. Comme dans le cas du filtrage temporel pur, la compensation des mouvements transparents implique un compromis entre le niveau du débruitage et la conservation des contrastes. Ce phénomène vient du fait que la compensation des mouvements transparents construit des images stationnaires temporellement, mais qui sont plus bruitées que les images originelles.

Le filtre hybride offre un compromis différent : en ne compensant les mouvements transparents qu'en certains pixels (ceux où les deux couches sont hétérogènes), on obtient un débruitage plus efficace, tout en conservant pour une large part les contrastes.

Les tests que nous avons effectués sur le cas particulier du filtre spatio-temporel adaptatif bilatéral étendu confirment ces enseignements. Ils montrent également que le cas du filtrage avec compensation des mouvements transparents peut constituer un compromis envisageable, ce qui n'était pas le cas pour les filtres temporels purs. La réduction de bruit est certes légèrement plus faible que pour les deux autres alternatives (pas de compensation, ou filtre hybride). Comme d'autre part les contrastes sont parfaitement conservés, le compromis permis par le filtre FSTCM est tout à fait intéressant.

Conclusion générale et perspectives

Dans cette étude, nous nous sommes intéressés à la problématique de l'estimation du mouvement des organes dans des séquences d'images acquises par rayons X. Une mesure fiable de ces mouvements permet de nombreuses applications, dont la stabilisation des séquences d'images (dans le cadre de l'angiographie soustraite par exemple), la compression efficace pour l'archivage ou l'extraction de métriques permettant de quantifier certaines insuffisances cardiaques. Dans le cadre de cette thèse, nous avons exploré comme application l'apport du mouvement pour les filtres réducteurs de bruit.

L'estimation des mouvements dans des séquences d'images acquises par rayons X présente un certain nombre de spécificités. Les images considérées sont relativement bruitées et peu contrastées. Surtout, leur formation résulte de l'atténuation du rayonnement lors de sa traversée de l'anatomie du patient, ce qui induit un effet de transparence qu'il est important de modéliser explicitement.

D'autre part, les images considérées permettent de faire des hypothèses simplificatrices portant sur la régularité des mouvements anatomiques et le nombre de couches transparentes pouvant se superposer en une région donnée de l'image, de sorte que le difficile problème initial peut être formulé de manière appréhendable.

Synthèse des travaux effectués

Estimation robuste des mouvements en situation de bitransparence. Des différentes voies existantes pour estimer les mouvements en situation de transparence, il apparaît que seules celles s'appuyant sur l'équation de contrainte du mouvement transparent sont applicables aux types de mouvements que nous considérons. Les deux autres alternatives ne s'y prêtent pas : la mesure des mouvements avec des estimateurs classiques ne peut pas appréhender la transparence de manière générale, et les approches fréquentielles nécessitent une constance temporelle des mouvements sur un horizon trop large.

Nous avons développé dans un premier temps une méthode d'estimation des mouvements anatomiques en situation de bi-transparence, c'est-à-dire lorsque deux couches exactement sont présentes sur toute l'image. Afin que l'approche soit aussi robuste au bruit et aux faibles contrastes que possible, nous avons contraint le problème en tenant compte du fait que les mouvements anatomiques étaient réguliers : nous avons postulé que les mouvements dans l'image suivaient un modèle de mouvement polynomial $2D$ (en pratique, un modèle affine).

Nous avons sur la base de cette hypothèse construit un schéma d'estimation paramétrique robuste dans un formalisme de Gauss-Newton. Nous avons développé une méthode d'initialisation en deux temps particulièrement efficace, et avons étudié le comportement de

la méthode globale à la fois sur des données synthétiques et des données réelles. Les résultats obtenus sont très satisfaisants, ce qui montre que les hypothèses effectuées sont raisonnables.

Estimation des mouvements dans le cas général. Nous avons alors considéré le cas général d'une séquence contenant un nombre quelconque de couches. L'observation de nombreuses séquences cliniques réelles nous a convaincus que, quel que soit le nombre de couches qu'elles contiennent globalement, les superpositions de plus de deux couches sont rares. Nous avons donc fait l'hypothèse de la bi-transparence distribuée : les séquences d'images considérées peuvent être segmentées en régions bi-transparentes.

Une première approche se place dans le cas où la segmentation en couches est connue (en pratique, donnée par l'utilisateur). Un schéma proche de celui utilisé en situation de bi-transparence simple permet d'estimer simultanément le modèle de mouvement de toutes les couches entrant en jeu dans la séquence. Seul le mécanisme d'initialisation a dû être reformulé : il consiste à initialiser les modèles des mouvements des couches à partir des modèles estimés sur chaque région bi-transparente.

Nous avons ensuite proposé une méthode jointe d'estimation des mouvements transparents et de segmentation des couches ne nécessitant aucun a-priori. Elle se base sur la minimisation itérative d'une fonction globale issue d'un formalisme markovien. L'initialisation par mise en correspondance par blocs, puis l'application d'une transformée de Hough, permettent d'obtenir une première estimation des modèles de mouvement, mais aussi de la segmentation et du nombre total de couches.

Ces deux méthodes ont été testées à travers des expériences sur données synthétiques, et surtout sur de nombreuses séquences cliniques, avec succès. Ceci valide en particulier la pertinence de l'hypothèse de bi-transparence distribuée. Les deux approches se sont révélées performantes dans leurs conditions idéales d'utilisation, et de plus elles sont peu perturbées par des transparences triples et la présence d'outils interventionnels.

Compensation des mouvements transparents. Dans un deuxième temps de nos travaux, nous nous sommes concentrés sur l'exploitation des modèles de mouvement estimés pour le débruitage. Le phénomène de transparence impose le développement d'une méthode spécifique de compensation des mouvements transparents, qui peut se déduire de l'équation de contrainte du mouvement transparent.

Toutefois, les images recalées par cette méthode sont entachées d'un niveau de bruit supérieur à celui des images exploitées par une approche sans compensation de mouvements transparents. Par conséquent, un filtre avec compensation des mouvements transparents (que ce soit dans un cadre temporel ou spatio-temporel) se traduit par une conservation bien meilleure des contrastes, mais par des performances globales plus faibles en terme de réduction du niveau de bruit que le même filtre sans compensation de mouvements transparents. Dans le cas particulier des filtres temporels, ce type d'approche se heurte même à une limite asymptotique de débruitage relativement élevée.

C'est pourquoi nous avons également proposé une approche hybride cherchant à ne compenser les mouvements transparents que sur les régions où les deux couches sont hétérogènes. Ceci permet d'y conserver les contrastes tout en filtrant de manière marquée le reste de l'image.

Les expériences effectuées sur des séquences synthétiques et cliniques ont montré que l'ap-

proche hybride conduit à des résultats plus intéressants que la simple compensation des mouvements transparents dans le cas du filtrage temporel pur. Dans le cas de filtrage spatio-temporel, les filtres avec compensation des mouvements transparents et les versions hybrides permettent des compromis “conservation des contrastes contre réduction du bruit” différents mais également intéressants.

Perspectives

Nous pouvons imaginer différents développements prometteurs des estimateurs de mouvements anatomiques en situation de transparence d’une part, et des filtrages du bruit avec prise en compte des mouvements transparents d’autre part.

Amélioration de l’estimation du mouvement et de la segmentation des couches en situation de bi-transparence distribuée. Bien que nous ayons démontré sur de nombreux exemples synthétiques, ainsi que sur une large base de séquences cliniques, que les estimations de mouvements anatomiques obtenues sont relativement précises, elles peuvent probablement encore être améliorées. Nous ne pensons pas qu’il soit pour cela nécessaire de remettre en cause la formalisation du problème que nous avons proposé, puisque les hypothèses de modèle affine du mouvement et de bi-transparence distribuée ont montré leur pertinence, et leur capacité à contraindre un problème difficile.

Une amélioration consisterait à affiner la segmentation spatiale en diminuant la taille des blocs considérés au cours de l’estimation. D’autre part, la minimisation de la fonction globale issue du formalisme markovien peut être remise en cause, par exemple en utilisant la méthode du *graph-cut*. Celle-ci permet une résolution optimale et rapide de problèmes de minimisation. C’est une approche bien maîtrisée dans le cas de segmentations à deux classes, mais dont la traduction dans notre cas particulier n’est pas directe puisque le nombre de couches entrant globalement en jeu peut être supérieur à deux et le nombre de valeurs différentes d’étiquetage (ou classes) peut devenir important pour ce qui est des régions bi-transparentes.

Détection des estimations erronées. Pour exploiter les algorithmes proposés en situation réelle, un module de détection des estimations erronées serait nécessaire. En effet, lorsque l’hypothèse de conservation temporelle du mouvement n’est pas du tout vérifiée (ce qui est le cas dans les phases de transition systole-diastole et diastole-systole), les mouvements estimés sont erronés. S’appuyer sur les mouvements calculés à ces instants risque de perturber l’application considérée. Ce phénomène a en particulier des conséquences importantes pour l’application majeure que nous considérons : le débruitage. Les filtres considérés ayant une structure récursive, une mauvaise estimation à un instant donné a des implications sur plusieurs des images suivantes.

Cette dépendance à l’hypothèse de conservation temporelle du mouvement est intrinsèque aux approches s’appuyant sur (2.39), auxquelles nous ne voyons pas d’alternatives pour estimer les mouvements transparents. L’occurrence d’estimations erronées est donc inévitable quelle que soit la mise en œuvre particulière de l’estimation, de sorte qu’il faudra les détecter

- soit de manière interne à l'estimation, en remarquant que le modèle de mouvement estimé ne correspond à aucune réalité physique (et qu'en particulier le résidu correspondant reste important) ;
- soit en utilisant des informations exogènes sur le rythme cardiaque pour identifier les instants critiques. Un accès à l'électrocardiogramme permettrait par exemple de repérer les dates susceptibles d'induire de mauvaises estimations des mouvements.

Une fois cette détection effectuée, il est envisageable

- d'estimer le mouvement cardiaque d'une autre façon, par exemple par une approche mono-couche, moins fiable mais ne faisant pas d'hypothèse de constance temporelle du mouvement ;
- d'interpoler la valeur manquante en s'appuyant sur l'historique des modèles de mouvement estimés (et sur un *a priori* sur le rythme cardiaque) ;
- d'adapter l'application pour qu'elle ne tienne pas compte de ces valeurs peu fiables.

Introduction d'un lissage temporel. Une piste d'amélioration des résultats est l'introduction d'un lissage temporel. Nous nous sommes concentrés sur l'estimation de mouvements transparents sur trois images consécutives, et n'avons pas tiré profit de la redondance des modèles de mouvement estimés (qui varient de manière régulière dans le temps) ou de la segmentation le long de la séquence.

Cette extension n'est pas triviale si l'on veut garder la robustesse de la méthode d'initialisation que nous avons développée, et la souplesse de la segmentation automatique (qui s'adapte à des éléments nouveaux). Ainsi, se contenter d'initialiser les modèles de mouvement et la segmentation par les résultats de l'instant précédent *dégrade* la qualité des résultats. C'est pourquoi nous pensons qu'il faut

- dérouler l'estimation en ignorant les résultats aux instants précédents dans un premier temps ;
- les lisser en s'appuyant sur l'historique des mouvements anatomiques estimés dans un deuxième temps. On pourra avoir recours à des *a priori* anatomiques à ce stade.

Une alternative peut consister à initialiser seulement la segmentation de l'image en couches à l'aide de la segmentation de l'instant précédent, déplacée par les mouvements correspondant aux modèles estimés.

Implantation en temps réel. Un autre aspect requis avant une utilisation clinique sera le développement d'une version temps réel de la méthode d'estimation des mouvements transparents. Elle est pour l'heure 3000 fois trop lente.

Une accélération de quelques ordres de grandeur est envisageable, sans changer l'architecture de l'estimateur

- d'abord parce qu'aucune technique spécialement coûteuse en temps de calcul n'a été introduite ;
- ensuite parce que ni le code, ni les réglages des paramètres n'ont cherché à rendre l'exécution du programme aussi rapide que possible⁶.

⁶Il est en particulier probablement possible de gagner beaucoup de temps en implantant mieux le calcul du facteur C de la fonction de Tukey (qui implique actuellement le calcul coûteux d'un médian sur la grille de l'image), d'accélérer les calculs de la fonction de Tukey en tabulant ses valeurs, de n'effectuer que trois itérations de la minimisation de la fonctionnelle markovienne, de ne pas effectuer l'estimation paramétrique au niveau 0 de la pyramide de résolution, et d'utiliser du matériel dédié pour calculer efficacement la mise

Relation avec le suivi d'outils interventionnels. Une autre perspective concerne le calcul du mouvement des outils interventionnels. Nous avons déjà expliqué que nous croyons qu'il faut développer des méthodes spécifiques pour les suivre de manière fiable : ils forment généralement des régions très effilées de sorte qu'une approche générale ne pourra pas en estimer le mouvement de façon suffisamment fiable. En outre, leurs mouvements plus saccadés que ceux des organes les rendent très sensibles à l'hypothèse de constance temporelle des mouvements.

Ces méthodes spécifiques de suivi pourront interagir avec les estimateurs de mouvements transparents anatomiques proposés dans cette étude :

- l'information du mouvement des organes sera utile à l'algorithme de suivi, en lui indiquant dans quelle direction rechercher l'outil ;
- la donnée des positions des outils permettra de traiter séparément leur suivi et l'estimation du mouvement des organes.

Le fait que la méthode d'estimation des mouvements anatomiques soit éventuellement capable parfois de capter les mouvements des outils est bien sûr un signe de la qualité de cet algorithme, mais il induit aussi des interactions parfois perturbatrices (notamment quand les mouvements des outils saccadés). Il sera donc appréciable de découpler ces deux aspects.

Développement d'autres applications. Les modèles de mouvement des organes estimés sont appelés à être utilisés dans diverses applications. Il peut s'agir d'aider au suivi des outils interventionnels, de quantifier des mouvements anatomiques ou de permettre une compression plus efficace par exemple. Ces applications sont encore à développer.

Utilisation pratique du débruitage. Enfin, des travaux majeurs restent à effectuer en ce qui concerne le débruitage pour valoriser, et le cas échéant utiliser de manière pratique la compensation hybride du mouvement transparent. Il s'agit d'un domaine de recherche très vaste, qui implique la comparaison avec les nombreuses méthodes de filtrage proposées dans la littérature et leur confrontation avec les besoins des cliniciens.

Annexes

Chapitre A

Séparation des couches transparentes sur images médicales

Nous avons implanté et testé les trois méthodes automatiques les plus matures de séparation de couches sur des séquences d'images cliniques. Il s'agit de :

- la stabilisation de la séquence par rapport à une couche, suivie du calcul du médian de la séquence stabilisée, que nous dénommerons par la suite *méthode du médian direct* ;
- la stabilisation de la séquence par rapport à une couche, suivie du calcul du médian des gradients horizontaux et verticaux des images stabilisées, et enfin de la reconstruction de la couche stabilisée. Cette dernière étape calcule à partir des deux images de gradient spatial de la couche stabilisée son laplacien, qui est ensuite inversé dans l'espace de Fourier [Weiss 01]. Nous parlerons de la *méthode du médian sur les gradients* ;
- la méthode d'*échange d'information des couches*, qui s'applique aux images transparentes dont le mouvement d'une des deux couches peut être stabilisé [Sarel 04].

Les deux dernières approches sont présentées plus en détail dans le chapitre 2. Nous commençons cette annexe par l'étude des deux premières méthodes citées avant de revenir sur la troisième.

A.1 Médian temporel sur séquences stabilisées

A.1.1 Tests synthétiques

Nous comparons les performances du filtre médian temporel direct et du médian temporel appliqué sur les gradients spatiaux sur des séquences synthétiques de 9 images composées de deux couches. L'une d'entre elles est fixe et l'autre en translation constante le long de la séquence.

La figure A.1 contient les résultats obtenus sur une séquence construite à partir d'images vidéos. Elle met en évidence le fait que le filtre médian direct permet une séparation grossière de la couche fixe, mais qu'un fantôme important de l'autre couche y est présent. La méthode du médian appliqué aux gradients permet une séparation avec beaucoup moins d'artefacts, même si la comparaison avec la vérité-terrain révèle qu'un fantôme plus léger de la deuxième couche est encore présent.

Les mêmes conclusions peuvent être tirées de la Fig.A.2, qui présente les résultats obtenus sur une séquence d'images médicales synthétiques réalistes (générée suivant la méthode de la section 1.3.2). On y observe en particulier un fantôme correspondant au disque de la colonne vertébrale à gauche de l'image, qui est plus net dans le cas d'une séparation des couches par le médian direct que par le médian appliqué aux gradients.



FIG. A.1: Séparation des couches transparentes sur une séquence synthétique de 9 images construite à partir d'images vidéo. En haut, les images 0, 4 et 8 de la séquence. En bas : les résultats du traitement à gauche pour le médian temporel direct, au centre pour le médian appliqué aux gradients, et à droite la vérité-terrain.

A.1.2 Influence de la qualité de la stabilisation

Ces deux méthodes supposent qu'une couche est immobile, ce qui implique donc que les mouvements de la couche d'intérêt ont été calculés et compensés (à l'exception du cas de la couche statique, sur laquelle nous revenons plus bas). Dans la pratique, ce recalage est imparfait du fait des erreurs d'estimation de mouvement commises.

Nous étudions leur influence à la figure A.3, où nous perturbons la couche stabilisée de la Fig.A.1 à chaque instant par un déplacement aléatoire d'amplitude maximale 1, 2 ou 5 pixels. Nous notons une dégradation très rapide de la qualité de la séparation, qui semble déjà inacceptable quand la stabilisation n'est exacte qu'à 2 pixels près.

Or, il faut considérer une dizaine d'images au moins pour que le médian permette une reconstruction intéressante de la couche stabilisée. À cet horizon temporel, les modèles de mouvement temporel ne peuvent pas être estimés par les méthodes présentées dans le corps

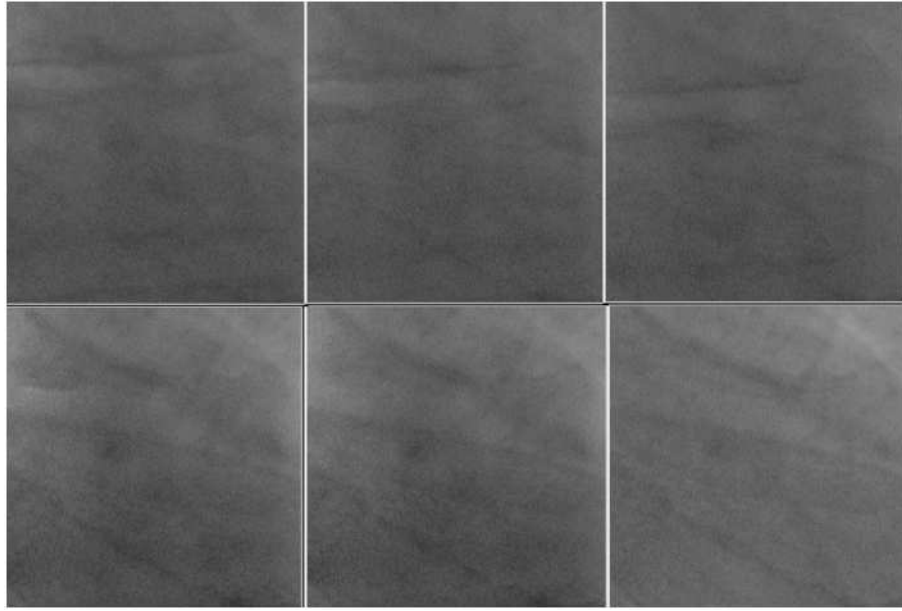


FIG. A.2: Séparation des couches transparentes sur une séquence synthétique de 9 images cliniques. En haut, les images 0, 4 et 8 de la séquence. En bas : résultats à gauche pour le médian temporel direct, au centre pour le médian appliqué aux gradients, et à droite la vérité-terrain.

du document¹ avec une précision suffisante pour permettre une séparation acceptable des couches.

Ces méthodes de séparation de couches **ne sont donc applicables** dans le cadre de notre problématique **qu'aux couches immobiles**. La plupart des séquences d'images fluoroscopiques réelles faisant intervenir un fond statique, ce cas particulier reste néanmoins intéressant.

A.1.3 Séparation de la couche statique

Nous appliquons ces deux schémas de séparation de la couche statique à la séquence fluoroscopique réelle numéro 10 (qui compte 90 images), dont nous extrayons deux images à la figure A.4. Elles sont prises aux deux extrémités du cycle respiratoire, de sorte que l'amplitude du mouvement du diaphragme y est visible. Une séparation idéale, au sens anatomique du terme, permettrait de reconstruire le fond statique : côtes, colonne vertébrale et tissus statiques.

Le résultat de la séparation de la couche statique suivant les deux méthodes considérées est montré à la figure A.5. Les images obtenues sont parfaitement débruitées (chaque pixel étant le médian de 90 pixels bruités!). Les séparations résultantes diffèrent du résultat idéal puisque le cœur et le diaphragme n'y sont que partiellement visibles.

¹Qui permettent une estimation des mouvement avec une précision moyenne de 0.5 à 2 pixels *entre deux images successives* suivant les configurations, et ce en supposant que l'hypothèse de constance temporelle du mouvement entre deux intervalles de temps successifs vérifiée.



FIG. A.3: Séparation des couches transparentes par la méthode du médian appliqué aux gradients, sur une séquence synthétique de 9 images construite à partir d'images vidéo. La qualité de la compensation du mouvement de la couche stabilisée varie : elle est parfaite en haut à gauche, exacte à 1 pixel près en haut à droite, à 2 pixels près en bas à gauche et à 5 pixels près en bas à droite.

Elles sont cela dit logiques au vu des méthodes utilisées. En particulier, la frontière du cœur en haut à gauche de l'image est quasi-immobile, et de ce fait impossible à séparer. Par contre, sa frontière en haut à droite a bien été effacée par la méthode du médian appliqué au gradient, mais pas par la méthode du médian direct. De même, l'image du diaphragme est plus atténuée par le médian appliqué au gradient que par la méthode directe.

Ceci montre la supériorité de la méthode du médian appliquée au gradient, et son applicabilité pour extraire un fond statique visuellement intéressant.

Nous avons représenté à la figure A.6 la séquence résiduelle (c'est-à-dire la séquence originale privée du fond calculé par la méthode du médian appliqué aux gradients). Nous observons un résultat qui ne correspond bien sûr pas à la séquence résiduelle idéale (qui serait composée du cœur et du diaphragme seulement) puisqu'une partie du cœur et du diaphragme ont été attribuées au fond statique.

Cela ne serait pas gênant outre mesure si la séquence résiduelle suivait la physique de la transparence additive, mais cela ne semble pas être le cas. En particulier, le diaphragme mobile, loin d'avoir un niveau de gris constant, est sombre lors de l'inspiration et clair lors de l'expiration.

Ces deux méthodes de séparation de couches permettent donc d'obtenir des résultats qua-

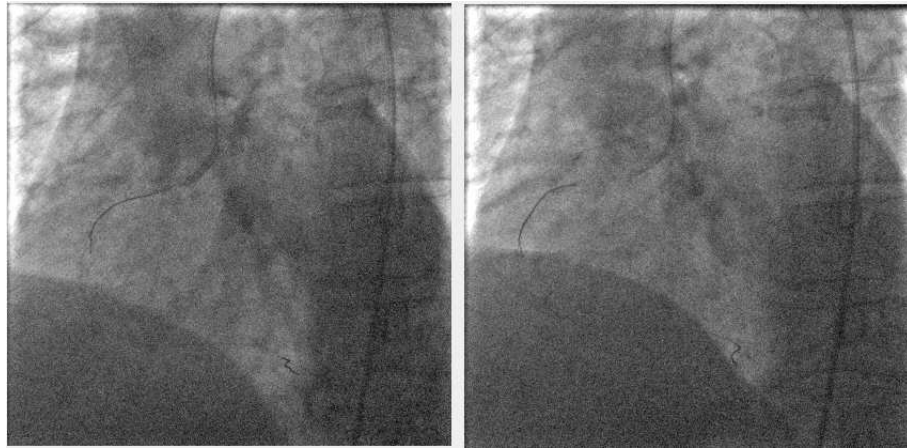


FIG. A.4: Deux images de la séquence fluoroscopique réelle, prises aux deux extrémités du cycle respiratoire.

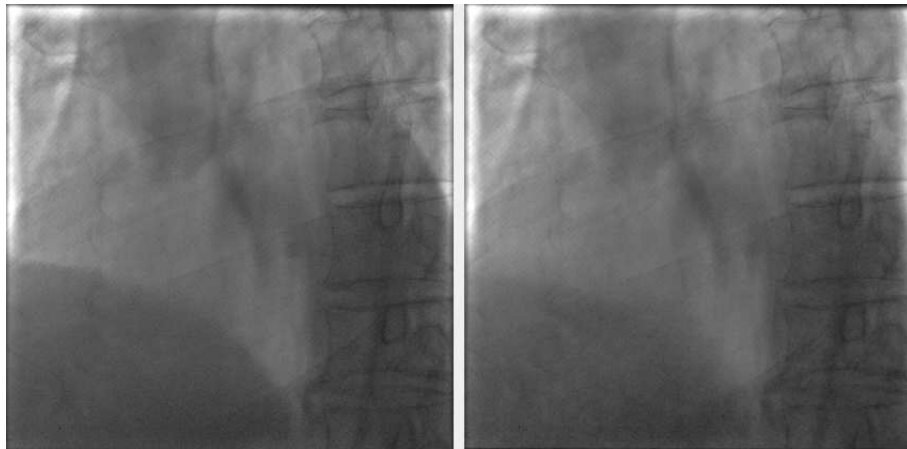


FIG. A.5: Séparation de la couche statique suivant les méthodes du médian direct (à gauche) et du médian appliqué aux gradients (à droite).

litativement intéressants, mais il est par contre illusoire de les inclure dans un schéma de compensation du mouvement transparent qui se doit de respecter la physique du problème.

A.2 Méthode d'échange d'informations au niveau des couches

A.2.1 Approche globale

Nous avons alors essayé d'enrichir la méthode du médian appliqué au gradient en utilisant le mécanisme d'échange d'informations des couches proposé par Sarel et Irani, et décrit dans le chapitre 2. Cette méthode, originalement développée pour séparer des images

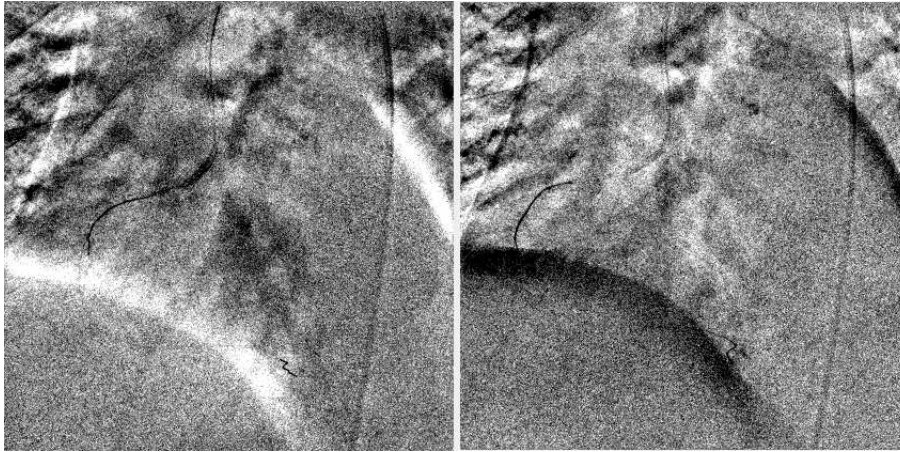


FIG. A.6: Séquence originale privée de la couche statique construite suivant la méthode du médiant direct, aux dates 47 et 85.

acquises avec des filtres polarisés, s'applique à la séparation de couches sur séquences d'images transparentes. Les deux images considérées sont alors le fond statique extrait par la méthode du médiant appliqué au gradient, et une image résiduelle de la séquence. Cette méthode de séparation de couches permet, sur la base de mesures de corrélation locales, de transmettre l'information correspondant au fantôme d'une des deux couches à l'aussi image extraite, jusqu'à ce que les deux couches séparées soient autant indépendantes que possible.

La méthode est appliquée dans un premier temps à l'échelle de toute l'image, avant d'être utilisée sur des blocs de plus en plus petits. Pour des raisons de stabilité sur lesquelles nous allons revenir, nous l'appliquons dans un premier temps sur l'image dans son intégralité. Nous comparons à la figure A.7 la couche statique séparée obtenue par la méthode du médiant appliqué aux gradients, et celle issue de cette méthode d'échange d'informations. Nous ne notons aucune différence significative, et de fait le transfert d'informations a été très faible.

A.2.2 Approche locale

Nous avons alors utilisé la version locale de l'échange d'informations entre couches, en réduisant progressivement la taille des blocs sur lesquels s'effectue le transfert d'information jusqu'à atteindre la taille 15×15 (comme conseillé dans [Sarel 04]). Cependant, nous avons constaté l'apparition d'artefacts par blocs que nous mettons en évidence sur des images synthétiques construites à partir d'images vidéo à la figure A.8. Ces artefacts ne disparaissent pas en utilisant des blocs qui se chevauchent.

Il semble en fait que la taille des blocs doit être choisie manuellement pour être en accord avec l'échelle des détails présents sur l'image, sans quoi le résultat devient instable. Il n'est donc pas envisageable d'utiliser cette approche sur les images peu contrastées et très bruitées que nous considérons.

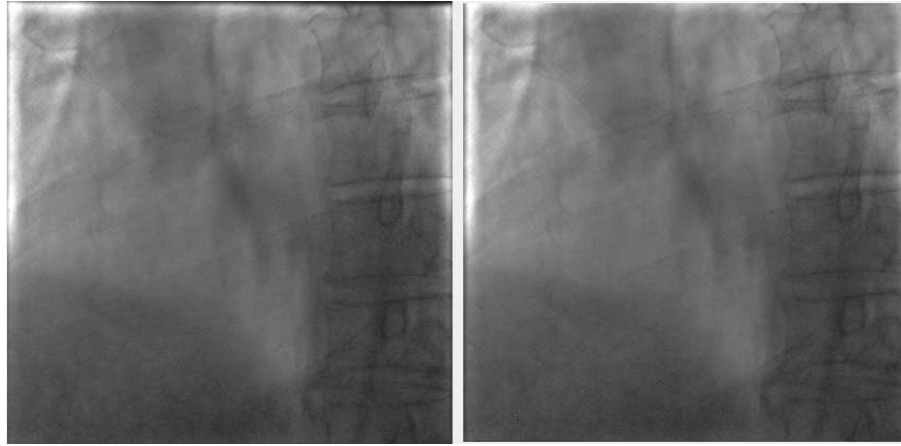


FIG. A.7: Séparation de la couche statique suivant la méthode du médian appliqué aux gradients seuls (à gauche), et suivant la méthode du médian appliqué au gradient suivi de l'échange d'informations des couches (à droite).

A.3 Conclusion

Nous avons étudié dans cette annexe l'applicabilité des méthodes de séparation des couches transparentes aux images fluorocopiques. Nous avons conclu que l'utilisation du médian temporel sur les séquences stabilisées donnait des résultats qualitativement intéressants, et qui étaient meilleurs lorsque le médian était appliqué sur les gradient spatiaux d'intensité.

Cependant, ces méthodes ne sont applicables en pratique qu'à la couche statique, car elles sont très sensibles à la qualité de la compensation du mouvement de la couche considérée. En outre, si la couche séparée fournit une bonne approximation visuelle de la couche réelle, elle ne vérifie pas précisément la physique de la transparence additive, de sorte que la séquence d'images résiduelles n'est pas exploitable.

Nous avons alors essayé d'améliorer la séparation par la méthode d'échange d'informations, mais sans succès. Quand elle est utilisée au niveau de l'image, le transfert d'information est si faible que les couches extraites ne sont quasiment pas modifiées, alors que quand elle est appliquée de manière plus locale, des problèmes de stabilité apparaissent.

Au final donc, aucune des méthodes actuelles de séparation des couches n'est applicable à notre problématique, que ce soit pour estimer les mouvements transparents ou pour débiter les images transparentes dans le sens du mouvement. Soulignons en outre que les méthodes présentées impliquent un volume de calcul important puisque le médian doit être pris sur une dizaine d'images successives au moins pour fournir des résultats intéressants.



FIG. A.8: Deux images transparentes mixtes, avec leur séparation par la méthode de l'échange d'informations.

Chapitre B

Minimisation du critère de bitransparence linéarisé

Nous développons dans cette annexe les calculs menant à l'estimation de mouvements transparents par un schéma de Gauss-Newton, que ce soit en situation de bi-transparence ou de bi-transparence distribuée. Nous commençons par rappeler les équations sur lesquelles nous nous appuyons, et par introduire les notations englobant bi-transparence et bi-transparence distribuée dans un même formalisme. Nous mènerons ensuite les calculs de minimisation de l'énergie en nous appuyant sur un résidu linéarisé au premier ordre, puis au second ordre.

B.1 Estimation de mouvements paramétriques par la technique de Gauss-Newton

B.1.1 L'équation de contrainte du mouvement transparent impliquant des modèles paramétriques

Nous partons de l'équation de contrainte du mouvement transparent. Si la séquence d'images I est composée de deux couches transparentes se déplaçant avec les vitesses \mathbf{w}_1 et \mathbf{w}_2 respectivement, alors :

$$\begin{aligned} r(\mathbf{p}, \mathbf{w}_1, \mathbf{w}_2) &= I(\mathbf{p}, t+1) - I(\mathbf{p} + \mathbf{w}_1, t) \\ &\quad - I(\mathbf{p} + \mathbf{w}_2, t) + I(\mathbf{p} + \mathbf{w}_1 + \mathbf{w}_2, t-1) = 0 \end{aligned} \quad (\text{B.1})$$

Estimer les mouvements transparents sur toute l'image revient à minimiser :

$$J(\mathbf{w}_1, \mathbf{w}_2) = \sum_{\mathbf{p} \in \mathfrak{S}} \left(r(\mathbf{p}, \mathbf{w}_1(\mathbf{p}), \mathbf{w}_2(\mathbf{p})) \right)^2 \quad (\text{B.2})$$

où \mathbf{w}_1 et \mathbf{w}_2 représentent les champs de vitesses associés aux modèles paramétriques θ_1 et θ_2 des couches. Ils doivent être approximativement constants localement pour que cette équation soit valide (voir chapitre 2).

Comme nous adoptons en fait un estimateur robuste, la minimisation est réalisée en fait par une technique de moindre carrés pondérés itérés. Pour simplifier les notations, nous

restons pour les développements qui suivent au cas des moindres carrés (relation (B.2)), l'extension aux moindres carrés pondérés étant immédiate.

Les vecteurs de vitesse \mathbf{w}_i s'écrivent de la façon suivante :

$$\mathbf{w}_i = \mathbf{w}(\mathbf{p}, \theta_i) = \mathbf{B}(\mathbf{p})\theta_i \quad (\text{B.3})$$

où la matrice $\mathbf{B}(\mathbf{p})$ vaut (cas de l'utilisation des modèles affines complets) :

$$\mathbf{B}(\mathbf{p}) = \begin{bmatrix} 1 & x & y & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x & y \end{bmatrix} \quad (\text{B.4})$$

Ainsi, nous devons estimer les paramètres θ_i pour mesurer tous les mouvements de l'image. Évoquons dès maintenant une propriété de ce modèle affine, qui nous sera utile plus loin : \mathbf{w}_{θ_i} est linéaire en les différents paramètres de mouvement. En conséquence, une variation incrémentale des paramètres $\theta'_i = \theta_i + \delta\theta_i$ se traduit comme suit :

$$\mathbf{w}(\mathbf{p}, \theta_i + \delta\theta_i) = \mathbf{w}(\mathbf{p}, \theta_i) + \mathbf{w}(\mathbf{p}, \delta\theta_i) \quad (\text{B.5})$$

B.1.2 Un formalisme unique pour les deux configurations de transparence

Nous nous intéressons dans les chapitres 3, 4 et 5 à deux configurations différentes de transparence :

- Dans le chapitre 3, nous nous plaçons en situation de *bi-transparence*. Dans ce cas, il y a deux couches exactement dans toute l'image, présentes en tout pixel. Les vitesses $\mathbf{w}_1(\mathbf{p})$ et $\mathbf{w}_2(\mathbf{p})$ sont donc dépendantes des deux mêmes vecteurs de paramètres θ_1 et θ_2 en tout point \mathbf{p} de l'image.
- Dans les chapitres 4 et 5, nous traitons le cas de la *bi-transparence distribuée*. Il y a alors un nombre arbitraire de couches dans l'image, mais jamais plus de deux au même point. Cette annexe traite le cas où les étiquettes des modèles présents en chaque pixel sont connus ; il peut être soit donné par l'utilisateur (chapitre 4), soit calculé itérativement (chapitre 5).

Pour ne pas écrire des calculs hautement redondants, nous considérons dans cette annexe un formalisme unifié pour ces deux configurations.

Nous cherchons à estimer les différents modèles de mouvement θ_i présents dans l'image, en minimisant :

$$J(\Theta) = \sum_{\mathbf{p} \in \mathfrak{S}} r(\mathbf{p}, \mathbf{w}(\mathbf{p}, \theta_{e_{\mathbf{p}}(1)}), \mathbf{w}(\mathbf{p}, \theta_{e_{\mathbf{p}}(2)}), t + 1)^2 \quad (\text{B.6})$$

$$= \sum_{\mathbf{p} \in \mathfrak{S}} \left(I(\mathbf{p}, t + 1) - I(\mathbf{p} + \mathbf{w}(\mathbf{p}, \theta_{e_{\mathbf{p}}(1)}), t) \right. \\ \left. - I(\mathbf{p} + \mathbf{w}(\mathbf{p}, \theta_{e_{\mathbf{p}}(2)}), t) + I(\mathbf{p} + \mathbf{w}(\mathbf{p}, \theta_{e_{\mathbf{p}}(1)} + \mathbf{w}(\mathbf{p}, \theta_{e_{\mathbf{p}}(2)}), t - 1) \right)^2 \quad (\text{B.7})$$

où Θ est la concaténation des différents paramètres θ_i , $e_{\mathbf{p}}(1)$ et $e_{\mathbf{p}}(2)$ les étiquettes des modèles de mouvement présents en \mathbf{p} . Ces étiquettes sont supposées connues :

- elles valent respectivement 1 et 2 en tout point de l'image en situation de bitransparence globale ;

– elles sont issues de l'étape de segmentation en situation de bitransparence distribuée. Avec des notations allégées, nous avons :

$$J(\Theta) = \sum_{\mathbf{p} \in \mathfrak{S}} \left(I(\mathbf{p}, t+1) - I(\mathbf{p} + \mathbf{w}_{e_{\mathbf{p}}(1)}(\mathbf{p}), t) - I(\mathbf{p} + \mathbf{w}_{e_{\mathbf{p}}(2)}(\mathbf{p}), t) + I(\mathbf{p} + \mathbf{w}_{e_{\mathbf{p}}(1)}(\mathbf{p}) + \mathbf{w}_{e_{\mathbf{p}}(2)}(\mathbf{p}), t-1) \right)^2 \quad (\text{B.8})$$

B.1.3 Minimisation par la technique de Gauss-Newton

Le critère non linéaire (B.8) est difficile à minimiser directement. Nous choisissons de suivre un schéma incrémental de minimisation itérative de type Gauss-Newton pour estimer les paramètres de mouvement. Cela consiste à linéariser itérativement (B.8) autour des valeurs courantes des estimés, et à rechercher son minimum. Nous lui adjoignons une structure multi-résolution pour appréhender les larges déplacements.

Nous avons le choix de l'ordre de linéarisation du résidu. Les sections suivantes présentent deux variantes au premier et au second ordre respectivement. Le premier ordre est le choix le plus classique et il mène à une fonction dont le minimum exact se calcule explicitement. Le choix du second ordre s'est révélé en fait plus performant. Nous montrons en effet à travers les expériences décrites à la section 3.5 que la minimisation associée est certes plus onéreuse, mais qu'elle fournit des estimés plus fiables, ce qui réduit drastiquement le nombre de linéarisations successives nécessaires à l'objectif d'une précision suffisante.

B.2 Linéarisation au premier ordre

Le cas du premier ordre est la copie conforme des calculs détaillés pour la bitransparence simple en 3.2.1.1. Nous souhaitons minimiser J à partir d'une linéarisation de r au premier ordre par rapport autour de $\hat{\Theta}_k$. Nous le noterons par la suite simplement Θ pour alléger les notations. Considérons

$$J(\Theta + \delta\Theta) = \sum_{\mathbf{p} \in \mathfrak{S}} r^2(\mathbf{p}, \Theta + \delta\Theta) \quad (\text{B.9})$$

Or, l'expression de $r(\mathbf{p}, \Theta + \delta\Theta)$ est donnée d'après (B.1) par :

$$\begin{aligned} r(\mathbf{p}, \Theta + \delta\Theta) &= I(\mathbf{p}, t+1) - I(\mathbf{p} + \mathbf{w}_{\theta_{e_1(\mathbf{p})} + \delta\theta_{e_1(\mathbf{p})}}(\mathbf{p}), t) - I(\mathbf{p} + \mathbf{w}_{\theta_{e_2(\mathbf{p})} + \delta\theta_{e_2(\mathbf{p})}}(\mathbf{p}), t) \\ &\quad + I(\mathbf{p} + \mathbf{w}_{\theta_{e_1(\mathbf{p})} + \delta\theta_{e_1(\mathbf{p})}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_2(\mathbf{p})} + \delta\theta_{e_2(\mathbf{p})}}(\mathbf{p}), t-1) \quad (\text{B.10}) \\ &= I(\mathbf{p}, t+1) - I(\mathbf{p} + \mathbf{w}_{\theta_{e_1(\mathbf{p})}}(\mathbf{p}) + \mathbf{w}_{\delta\theta_{e_1(\mathbf{p})}}(\mathbf{p}), t) \\ &\quad - I(\mathbf{p} + \mathbf{w}_{\theta_{e_2(\mathbf{p})}}(\mathbf{p}) + \mathbf{w}_{\delta\theta_{e_2(\mathbf{p})}}(\mathbf{p}), t) \\ &\quad + I(\mathbf{p} + \mathbf{w}_{\theta_{e_1(\mathbf{p})}}(\mathbf{p}) + \mathbf{w}_{\delta\theta_{e_1(\mathbf{p})}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_2(\mathbf{p})}}(\mathbf{p}) + \mathbf{w}_{\delta\theta_{e_2(\mathbf{p})}}(\mathbf{p}), t-1) \quad (\text{B.11}) \end{aligned}$$

d'après la propriété de linéarité (B.5). En utilisant le développement de Taylor au premier ordre pour chaque terme, où ∇ désigne le gradient par rapport aux coordonnées spatiales

(x et y) de l'image, on obtient :

$$\begin{aligned}
r(\mathbf{p}, \Theta + \delta\Theta) &\simeq I(\mathbf{p}, t+1) - I(\mathbf{p} + \mathbf{w}_{\theta_{e_1}(\mathbf{p})}(\mathbf{p}), t) - \mathbf{w}_{\delta\theta_{e_1}(\mathbf{p})}^T \nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_1}(\mathbf{p})}(\mathbf{p}), t) \\
&\quad - I(\mathbf{p} + \mathbf{w}_{\theta_{e_2}(\mathbf{p})}(\mathbf{p}), t) - \mathbf{w}_{\delta\theta_{e_2}(\mathbf{p})}^T \nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_2}(\mathbf{p})}(\mathbf{p}), t) \\
&\quad + I(\mathbf{p} + \mathbf{w}_{\theta_{e_1}(\mathbf{p})}(\mathbf{p}) + \mathbf{w}_{\theta_{e_2}(\mathbf{p})}(\mathbf{p}), t-1) \\
&\quad + (\mathbf{w}_{\delta\theta_{e_1}(\mathbf{p})} + \mathbf{w}_{\delta\theta_{e_2}(\mathbf{p})})^T \nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_1}(\mathbf{p})}(\mathbf{p}) + \mathbf{w}_{\theta_{e_2}(\mathbf{p})}(\mathbf{p}), t-1) \quad (\text{B.12}) \\
&= r(\mathbf{p}, \Theta) \\
&\quad + (\nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_1}(\mathbf{p})}(\mathbf{p}) + \mathbf{w}_{\theta_{e_2}(\mathbf{p})}(\mathbf{p}), t-1)^T \\
&\quad \quad - \nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_1}(\mathbf{p})}(\mathbf{p}), t)^T) \mathbf{w}_{\delta\theta_{e_1}(\mathbf{p})}(\mathbf{p}) \\
&\quad + (\nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_1}(\mathbf{p})}(\mathbf{p}) + \mathbf{w}_{\theta_{e_2}(\mathbf{p})}(\mathbf{p}), t-1)^T \\
&\quad \quad - \nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_2}(\mathbf{p})}(\mathbf{p}), t)^T) \mathbf{w}_{\delta\theta_{e_2}(\mathbf{p})}(\mathbf{p}) \quad (\text{B.13})
\end{aligned}$$

$$\begin{aligned}
&= r(\mathbf{p}, \Theta) \\
&\quad + (\nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_1}(\mathbf{p})}(\mathbf{p}) + \mathbf{w}_{\theta_{e_2}(\mathbf{p})}(\mathbf{p}), t-1)^T \\
&\quad \quad - \nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_1}(\mathbf{p})}(\mathbf{p}), t)^T) \mathbf{B}(\mathbf{p}) \delta\theta_{e_1}(\mathbf{p}) \\
&\quad + (\nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_1}(\mathbf{p})}(\mathbf{p}) + \mathbf{w}_{\theta_{e_2}(\mathbf{p})}(\mathbf{p}), t-1)^T \\
&\quad \quad - \nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_2}(\mathbf{p})}(\mathbf{p}), t)^T) \mathbf{B}(\mathbf{p}) \delta\theta_{e_2}(\mathbf{p}) \quad (\text{B.14})
\end{aligned}$$

Ce que nous notons :

$$r(\mathbf{p}, \Theta + \delta\Theta) = r(\mathbf{p}, \Theta) + \nabla_{\theta} r(\mathbf{p}, \Theta)^T \delta\Theta \quad (\text{B.15})$$

L'énergie J à minimiser devient :

$$J_1(\delta\Theta) = \sum_{\mathbf{p} \in \mathfrak{S}} \left(r(\mathbf{p}, \Theta) + \delta\Theta^T \nabla_{\theta} r(\mathbf{p}, \Theta) \right)^2 \quad (\text{B.16})$$

$$\begin{aligned}
&= \sum_{\mathbf{p} \in \mathfrak{S}} r(\mathbf{p}, \Theta)^2 + 2r(\mathbf{p}, \Theta) \nabla_{\theta} r(\mathbf{p}, \Theta)^T \delta\Theta \\
&\quad + \delta\Theta^T \nabla_{\theta} r(\mathbf{p}, \Theta) \nabla_{\theta} r(\mathbf{p}, \Theta)^T \delta\Theta \quad (\text{B.17})
\end{aligned}$$

La solution de ce problème classique des moindres carrés est :

$$\delta\hat{\Theta} = - \left(\sum_{\mathbf{p} \in \mathfrak{S}} \nabla_{\theta} r(\mathbf{p}, \hat{\Theta}_k)^T \nabla_{\theta} r(\mathbf{p}, \hat{\Theta}_k) \right)^{-1} \cdot \left(\sum_{\mathbf{p} \in \mathfrak{S}} r(\mathbf{p}, \hat{\Theta}_k) \nabla_{\theta} r(\mathbf{p}, \hat{\Theta}_k) \right) \quad (\text{B.18})$$

$$= \mathbf{M}^{-1} \mathbf{A} \quad (\text{B.19})$$

On peut donc calculer $\delta\hat{\Theta}$ à l'aide d'une inversion matricielle.

B.2.1 Algorithme de minimisation associé à l'équation de contrainte du mouvement avec transparence linéarisée au premier ordre

Le procédé peut être résumé par l'algorithme 6.

Algorithme 6 Algorithme de minimisation de l'équation de contrainte du mouvement avec transparence linéarisée au premier ordre.

- Recalage des images à partir des modèles de mouvement estimés à l'étape précédente.

 - Calcul des gradients spatiaux de $I(\mathbf{p} + \mathbf{w}_{e_{\mathbf{p}}(1)}, t)$, $I(\mathbf{p} + \mathbf{w}_{e_{\mathbf{p}}(2)}, t)$ et $I(\mathbf{p} + \mathbf{w}_{e_{\mathbf{p}}(1)} + \mathbf{w}_{e_{\mathbf{p}}(2)}, t + 1)$, puis calcul de $\nabla_{\theta} r(\mathbf{p}, \Theta)$.

 - Itérativement (dans un cadre robuste) :
 1. Mise à jour des poids.
 2. Initialisation de la matrice \mathbf{M} et du vecteur \mathbf{A} du système à 0.
 3. Parcours de l'image : pour chacune des deux couches présentes, ajout de la contribution du pixel \mathbf{p} au système d'équations : $\nabla_{\theta} r(\mathbf{p}, \Theta)$ à \mathbf{M} et $r(\mathbf{p}, \Theta)\nabla_{\theta} r(\mathbf{p}, \Theta)$ à \mathbf{A} .
 4. Résolution du système $\mathbf{M}\delta\Theta = \mathbf{A}$ (par exemple par pivot de Newton).

 - Mise à jour des paramètres Θ par ajout de $\delta\hat{\Theta}$.
-

B.3 Minimisation du critère de la bitransparence linéarisée au second ordre

B.3.1 Développement limité du critère au second ordre

De la même façon que dans la partie précédente, nous linéarisons r à partir d'une linéarisation de r au premier ordre par rapport autour de $\hat{\Theta}_k$ (noté simplement Θ dans les calculs à venir). $\mathbf{H}(\mathbf{p}, t)$ désigne le hessien spatial au pixel \mathbf{p} et à la date t . Nous avons au second ordre :

$$\begin{aligned} & r(\mathbf{p}, \mathbf{w}(\mathbf{p}, \theta_{e_{\mathbf{p}}(1)} + \delta\theta_{e_{\mathbf{p}}(1)}), \mathbf{w}(\mathbf{p}, \theta_{e_{\mathbf{p}}(2)} + \delta\theta_{e_{\mathbf{p}}(2)})) \\ &= I(\mathbf{p}, t + 1) - I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}), t) - I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}) + \mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t) \\ &\quad + I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}) + \mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t - 1) \end{aligned} \quad (\text{B.20})$$

$$\begin{aligned} &= I(\mathbf{p}, t + 1) \\ &\quad - I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}), t) - \mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p})^T \nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}), t) \\ &\quad - \frac{1}{2} \mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p})^T \mathbf{H}(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}), t) \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) \\ &\quad - I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t) - \mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p})^T \nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t) \\ &\quad - \frac{1}{2} \mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p})^T \mathbf{H}(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t) \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}) \\ &\quad + I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t - 1) \\ &\quad + (\mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p})^T + \mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p})^T) \nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t - 1) \\ &\quad + \frac{1}{2} (\mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}))^T \mathbf{H}(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t - 1) \\ &\quad (\mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p})) \end{aligned} \quad (\text{B.21})$$

$$\begin{aligned} &= (I(\mathbf{p}, t + 1) - I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}), t) - I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t) \\ &\quad + I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t - 1)) \\ &\quad + \mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p})^T (\nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t - 1) - \nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}), t)) \\ &\quad + \mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p})^T (\nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t - 1) - \nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t)) \\ &\quad + \frac{1}{2} \mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p})^T (\mathbf{H}(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t - 1) \\ &\quad - \mathbf{H}(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}), t)) \mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) \\ &\quad + \frac{1}{2} \mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p})^T (\mathbf{H}(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t - 1) \\ &\quad - \mathbf{H}(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t)) \mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}) \\ &\quad + \frac{1}{2} \mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p})^T (\mathbf{H}(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t - 1) \\ &\quad + \mathbf{H}^T(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t - 1)) \mathbf{w}_{\delta\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}) \end{aligned} \quad (\text{B.22})$$

$$\begin{aligned}
& r(\mathbf{p}, \mathbf{w}(\mathbf{p}, \theta_{e_{\mathbf{p}}(1)} + \delta\theta_{e_{\mathbf{p}}(1)}), \mathbf{w}(\mathbf{p}, \theta_{e_{\mathbf{p}}(2)} + \delta\theta_{e_{\mathbf{p}}(2)})) \\
&= (I(\mathbf{p}, t+1) - I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}), t) - I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t) \\
&\quad + I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t-1)) \\
&\quad + (\nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t-1)^T - \nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}), t)^T) \mathbf{B}(\mathbf{p}) \delta\theta_{e_{\mathbf{p}}(1)} \\
&\quad + (\nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t-1)^T - \nabla I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t)^T) \mathbf{B}(\mathbf{p}) \delta\theta_{e_{\mathbf{p}}(2)} \\
&\quad + \frac{1}{2} \delta\theta_{e_{\mathbf{p}}(1)}^T \mathbf{B}(\mathbf{p})^T (\mathbf{H}(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t-1) \\
&\quad \quad - \mathbf{H}(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}), t)) \mathbf{B}(\mathbf{p}) \delta\theta_{e_{\mathbf{p}}(1)} \\
&\quad + \frac{1}{2} \delta\theta_{e_{\mathbf{p}}(2)}^T \mathbf{B}(\mathbf{p})^T (\mathbf{H}(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t-1) \\
&\quad \quad - \mathbf{H}(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t)) \mathbf{B}(\mathbf{p}) \delta\theta_{e_{\mathbf{p}}(2)} \\
&\quad + \frac{1}{2} \delta\theta_{e_{\mathbf{p}}(1)}^T \mathbf{B}(\mathbf{p})^T (\mathbf{H}(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t-1) \\
&\quad \quad + \mathbf{H}^T(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}(\mathbf{p}) + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}(\mathbf{p}), t-1)) \mathbf{B}(\mathbf{p}) \delta\theta_{e_{\mathbf{p}}(2)} \tag{B.23}
\end{aligned}$$

Ce que nous notons :

$$r(\mathbf{p}, \Theta + \delta\Theta) = r(\mathbf{p}, \Theta) + \nabla_{\Theta} r(\mathbf{p}, \Theta)^T \delta\Theta + \frac{1}{2} \delta\Theta^T H_{\Theta}(r(\mathbf{p}, \Theta)) \delta\Theta \tag{B.24}$$

Dans ces conditions, nous sommes en mesure de calculer analytiquement la fonction suivante, qu'il convient de minimiser par rapport à $\delta\Theta$:

$$J_2(\hat{\Theta}_k + \delta\Theta) = \sum_{\mathbf{p} \in \mathfrak{S}} (r(\mathbf{p}, \hat{\Theta}_k + \nabla_{\Theta} r(\mathbf{p}, \hat{\Theta}_k)^T \delta\Theta + \frac{1}{2} \delta\Theta^T H_{\Theta}(r(\mathbf{p}, \hat{\Theta}_k)) \delta\Theta) \tag{B.25}$$

Par souci de clarté, nous introduisons la notation :

$$K(\delta\Theta) = J_2(\hat{\Theta}_k, \delta\Theta) \tag{B.26}$$

Comme le montre (3.24), K est un polynôme de degré 4, que nous ne pouvons pas minimiser directement comme dans le cas du premier ordre, puisqu'une dérivation mènerait à un polynôme de degré 3 à annuler (dans un espace de dimension 12!). Nous avons donc recours à la méthode de minimisation du *gradient conjugué*, qui fonctionne en deux temps. Il s'agit de choisir une direction de descente permettant de se rapprocher du minimum, puis de fixer le pas de descente de sorte à minimiser (B.26) au mieux. Le procédé est itéré un certain nombre de fois, et est stoppé quand les estimés successifs sont très proches.

B.3.2 Calcul de la direction de descente

Le calcul du gradient de K est direct grâce à (B.24). Cependant, nous préférons ne pas le choisir comme direction de descente car une telle approche court le risque, en ne tenant pas compte de l'historique de la minimisation, d'explorer sans cesse un sous-espace de l'espace des paramètres.

Pour éviter cet écueil, nous lui préférons une variante des gradients conjugués afin de établir

la direction de descente $\mathbf{d}(\delta\Theta^k)$. La formule de Polak et Ribière [Polak 71] construit la direction de descente en fonction des minima successivement trouvés $\delta\Theta^{k-1}$ comme suit :

$$\mathbf{d}(\delta\Theta^k) = -\nabla_{\delta\Theta}K(\delta\Theta^k) + \gamma d(\delta\Theta^{k-1}) \quad (\text{B.27})$$

avec :

$$\gamma = \frac{\left(\nabla_{\delta\Theta}K(\delta\Theta^k) - \nabla_{\delta\Theta}K(\delta\Theta^{k-1})\right)^T \nabla_{\delta\Theta}K(\delta\Theta^k)}{\nabla_{\delta\Theta}K(\delta\Theta^k)^T \nabla_{\delta\Theta}K(\delta\Theta^k)} \quad (\text{B.28})$$

Cette méthode propose un compromis entre une descente suivant la plus forte pente, et une correction poussant à suivre une direction orthogonale à celle suivie précédemment.

B.3.3 Homogénéisation de la direction de descente

Cependant, ce type de minimisation fonctionne mal quand les inconnues sont de type différent. Le critère linéarisé K serait en effet bien plus sensible dans les directions correspondant aux paramètres affines du premier ordre que dans les directions correspondant aux translations. Nous aurions alors des gradients d'amplitude plus grande dans certaines directions que dans d'autres, ce qui conduirait à rechercher les minima dans des hyperplans particuliers de l'espace de recherche (en l'occurrence les paramètres dy premier ordre, ce qui conduit à négliger les termes de translation).

Le principe est alors de pondérer les différentes variables, de sorte que leurs gradients aient le même ordre de grandeur. Nous appelons le nouvel espace construit l'*espace homogénéisé*. Typiquement, lorsque nous considérons une fonction du type :

$$f(c, a) = \sum_x g(c + ax), \quad (\text{B.29})$$

il est préférable d'introduire la fonction :

$$h(c, b) = \sum_x g(c + bx/X) = f(c, a) \quad (\text{B.30})$$

avec X une valeur typique de x . Alors $b = aX$.

Une descente de gradient itérative sur cette fonction donne alors :

$$c^{i+1} = c^i - \alpha \frac{h}{\partial c}(c^i, b^i) \quad (\text{B.31})$$

$$= c^i - \alpha \frac{f}{\partial c}(c^i, a^i) \quad (\text{B.32})$$

$$b^{i+1} = b^i - \alpha \frac{h}{\partial b}(c^i, b^i) \quad (\text{B.33})$$

$$= b^i - \alpha \frac{\partial a}{\partial b} \frac{\partial f}{\partial a}(c^i, a^i) \quad (\text{B.34})$$

$$= b^i - \alpha \frac{1}{X} \frac{\partial f}{\partial a}(c^i, a^i). \quad (\text{B.35})$$

Cette dernière équation étant équivalente à :

$$a^{i+1} = a^i - \alpha \frac{1}{X^2} \frac{\partial f}{\partial a}(c^i, a^i) \quad (\text{B.36})$$

Ainsi, il est équivalent d'évoluer dans l'espace homogénéisé et de rester dans l'espace initial, en pondérant le gradient et le pas de descente par $1/X$.

Dans le cas de nos déplacements affines, nous attribuons à chaque paramètre a_i un poids p_{a_i} valant 1 s'il s'agit d'une translation, et D s'il s'agit d'un paramètre affine du premier ordre, avec D une distance en pixels typique de l'image considérée (par exemple, la demi-largeur de l'image). Notons alors $\nabla_{a_{ij}}K$ le vecteur de composantes $\left[\frac{1}{p_{a_i}} \frac{\partial J}{\partial a_i}\right]$

L'algorithme du gradient conjugué devient alors :

$$\mathbf{d}(\delta\Theta^k) = -\frac{1}{p_{a_i}} \cdot \nabla_{a_i}K(\delta\Theta^k) + \gamma \mathbf{d}(\delta\Theta^{k-1}) \quad (\text{B.37})$$

Avec :

$$\gamma = \frac{\left(\nabla_{a_i}K(\delta\Theta^k) - \nabla_{a_i}K(\delta\Theta^{k-1})\right)^T \nabla_{a_i}K(\delta\Theta^k)}{\nabla_{a_i}K(\delta\Theta^k)^T \nabla_{a_i}K(\delta\Theta^k)} \quad (\text{B.38})$$

B.3.4 Choix du pas

Une fois la direction de descente $\mathbf{d}(\delta\Theta^k)$ déterminée, il faut spécifier le pas α^k :

$$\delta\Theta^{k+1} = \delta\Theta^k + \alpha^k \mathbf{d}(\delta\Theta^k) \quad (\text{B.39})$$

Nous devons donc choisir à chaque itération α^{k+1} pour que le critère (B.8) soit minimal. Reprenons (B.24). Nous avons :

$$r(\mathbf{p}, \Theta + \delta\Theta) = r(\mathbf{p}, \Theta) + \delta\Theta^T \nabla_{\Theta} r(\mathbf{p}, \Theta) + \frac{1}{2} \delta\Theta^T H_{\Theta}(r(\mathbf{p}, \Theta)) \delta\Theta \quad (\text{B.40})$$

Injecter (B.39) dans (B.40) met en évidence que l'équation (B.40) est polynômiale d'ordre 2 en α^k .

Or, nous souhaitons minimiser :

$$J(\hat{\Theta}_k + \delta\Theta) = \sum_{\mathbf{p} \in \mathfrak{S}} r(\mathbf{p}, \hat{\Theta}_k + \delta\Theta)^2 \quad (\text{B.41})$$

qui est donc d'ordre 4 en α^k . Pour minimiser ce polynôme (en dimension 1), nous annulons sa dérivée. Il existe des formules exactes permettant de trouver les racines d'un polynôme d'ordre 3, mais elles sont numériquement moins bonnes (et plus longues à calculer) que d'annuler itérativement cette expression en utilisant la formule de Newton :

$$x_{n+1} = x_n - \frac{y(x_n)}{y'(x_n)} \quad (\text{B.42})$$

B.3.5 Algorithme de minimisation associé à l'équation contrainte du mouvement avec la transparence linéarisée au second ordre

Le procédé peut être résumé par l'algorithme 7.

Algorithme 7 Algorithme de minimisation associé à l'équation contrainte du mouvement avec la transparence linéarisée au second ordre.

- Recalage des images par les modèles de mouvement estimés à l'étape précédente.
- Calcul des gradients spatiaux et des dérivées secondes de $I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}}, t)$, $I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}, t)$ et $I(\mathbf{p} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(1)}} + \mathbf{w}_{\theta_{e_{\mathbf{p}}(2)}}, t + 1)$, puis de $\nabla_{\Theta} r(\mathbf{p}, \Theta)$ et de $H_{\Theta}(r(\mathbf{p}, \Theta))$.
- Dans un cadre robuste, remise des poids à jour puis

Pour les itérations $i = 1, 2, \dots$ des moindres carrés robustes,

Pour les itérations $k = 1, 2, \dots$ de la descente de gradient,

1. Parcours de l'image pour calculer en chaque point :
 - $\nabla K(\mathbf{p}, \delta\Theta^{k+1})$ par (B.24).
 - En utilisant la formule de Polak et Ribiere, et en prenant la précaution d'homogénéiser les paramètres, calcul de la direction de descente $\mathbf{d}(\Theta^k)$.
 2. Nouveau parcours de l'image pour calculer les coefficients du polynôme $K(\delta\Theta^k + \alpha^k \mathbf{d}(\Theta^k))$, de degré 4 en α^k .
 3. Annulation de la dérivée du polynôme par la formule de Newton. On pose alors $\delta\Theta^{k+1} = \delta\Theta^k + \alpha^k \mathbf{d}(\Theta^k)$.
- Mise à jour des paramètres Θ par ajout de $\delta\Theta^k$.
-

Chapitre C

Ajout de couches

Si les mécanismes de réduction du nombre de couches décrits en 5.3 sont faciles à mettre en œuvre, des méthodes plus complexes pour décider de l'ajout d'une couche doivent être envisagées.

Un indicateur pertinent s'avère être la valeur des poids issus de l'estimation robuste des modèles affines de mouvement. Si une couche physique n'est pas appréhendée par la segmentation courante, le résidu (2.39) n'y sera pas minimal, ce qui se traduira par des poids plus faibles sur les blocs en question qu'ailleurs. Si nous sommes en mesure de repérer ces blocs mal étiquetés, nous pouvons décider de créer une couche supplémentaire les représentant s'ils sont en nombre suffisant.

Cette approche pose néanmoins certaines questions.

- Comment détecter un bloc mal étiqueté ?
- Comment déterminer la couche du bloc qui est inappropriée ?
- Comment estimer le nouveau modèle de mouvement pour la couche introduite ?
- À quel moment placer ce contrôle ?

C.1 Détection de blocs mal étiquetés.

Que l'étiquetage d'un bloc s soit cohérent ou non avec les couches qui y sont réellement présentes, les pixels correspondant à des zones peu texturées se voient attribuer un poids proche de 1. Comme ils sont majoritaires dans la plupart des blocs, prendre en compte la moyenne des poids sur chaque bloc paraît délicat. C'est pourquoi nous optons pour un autre indicateur qui traduit une information plus pertinente : *nous comptons le nombre $N_o(s)$ de pixels du bloc dont le poids est inférieur à un certain seuil w_s* . Ces pixels seront appelés par la suite *outliers*.

Nous présentons sur la Fig.C.1 une carte des poids, où les pixels tendent vers le blanc quand leur poids est proche de 1, et vers le noir quand il est proche de 0. La séquence d'images considérée contient trois couches : l'une d'entre elles est présente sur l'ensemble de l'image, tandis que les deux autres sont placées en haut et en bas de l'image. Leurs frontières sont représentées en bleu et en rouge sur la carte des poids. Elles sont séparées par une zone mono-couche à laquelle nous ne nous intéressons pas pour l'heure.

Sur l'exemple présenté, les mouvements de la couche présente sur toute l'image et de

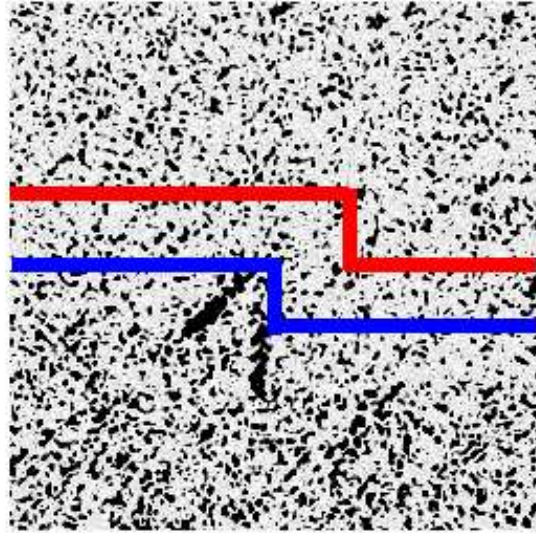


FIG. C.1: Carte des poids issus de l'estimation robuste des modèles de mouvement sur une séquence en situation de bi-transparence distribuée à trois couches (une couvrant toute l'image, les deux autres étant présentes sur chaque moitié haut et bas de l'image et délimitées par les lignes bleues et rouges). Les poids faibles tendent vers le noir et les poids forts vers le blanc.

celle du haut sont bien estimés, mais ce n'est pas le cas de celle du bas. Nous observons que dans la zone qui lui correspond les poids faibles sont plus denses. Cela dit, la différence visuelle n'est pas si évidente et on comprend bien que la moyenne des poids ne fournit pas un indicateur des plus fiables pour détecter cette configuration.

Nous indiquons sur Tab.C.1 le nombre de poids valant 0 des différents blocs, et remarquons que les zones mal étiquetées y apparaissent beaucoup plus nettement.

TAB. C.1: Nombre des poids nuls des différents blocs (rangés dans le tableau comme leur positionnement dans l'image)

98	69	107	94	126	89	102	92
105	143	116	102	119	86	102	109
98	134	114	105	105	99	122	99
74	121	120	92	80	69	102	154
153	177	173	310	106	74	91	126
185	206	222	285	119	155	82	116
194	283	164	192	231	299	191	181
191	195	228	198	176	222	188	197

Nous montrerons plus loin que le choix du seuil sur les poids w_s ne change pas la

discussion qui suit. Considérons-le pour l'heure fixé. La question est alors de choisir un seuil N_o^l sur le nombre d'outliers pour lequel un bloc sera considéré comme mal étiqueté. Du fait du bruit important, et de l'estimation paramétrique du mouvement qui n'est pas parfaite à convergence, des outliers sont aussi présents sur des blocs correctement étiquetés, et dont l'estimation du modèle du mouvement des couches correspondantes est bonne.

Or, le nombre typique d'outliers pour un bloc mal étiqueté peut dépendre de plusieurs facteurs : le niveau de bruit, la qualité de l'estimation du modèle de mouvement à ce point d'exécution de l'algorithme, l'anatomie imagée (qui est plus ou moins contrastée), etc. Nous décidons donc d'apprendre ce seuil *en ligne*.

Pour ce faire, nous calculons la valeur médiane N_o de $N_o(s)$ sur l'image, puis sa déviation médiane ΔN_o . Un bloc s est alors considéré comme mal étiqueté lorsque :

$$N_o(s) > N_o^l = N_o + \gamma \cdot \Delta N_o \quad (\text{C.1})$$

$$\text{avec } N_o = \underset{s \in S}{\text{med}} (N_o(s)) \quad (\text{C.2})$$

$$\text{et } \Delta N_o = \underset{s \in S}{\text{med}} (|N_o(s) - N_o|) \quad (\text{C.3})$$

Lorsque plus de m blocs sont détectés comme mal étiquetés, nous décidons d'introduire une nouvelle couche.

Sur l'exemple de la figure C.1, pour $w_s = 0$, $N_o = 120$ et $\Delta N_o = 33$. Nous présentons sur la figure C.6 les blocs détectés comme mal étiquetés pour différentes valeurs de γ . Nous remarquons que la dépendance à γ n'y est pas cruciale. Qu'il vale 1 ou 3, un nombre suffisant de blocs seront détectés pour commander l'ajout d'une couche.

Il reste trois paramètres à fixer :

- le seuil sur les poids w_s , qui définit les outliers.
- le paramètre γ qui fixe la limite sur le nombre d'outliers possibles dans un bloc correctement étiqueté.
- le nombre minimal m de blocs mal étiquetés pour décider d'ajouter une nouvelle couche.

Le nombre m peut être défini à partir de la taille minimale qu'une couche doit avoir pour être considérée comme pertinente. Nous avons choisi la valeur $m = 5$, qui est relativement élevée étant donné que nous ne cherchons à appréhender que le mouvement des organes.

Pour ce qui est du choix de w_s , il ne s'avère pas influent. Nous illustrons ce point sur le cas particulier de la figure C.1, en produisant à la figure C.7 les décomptes d'outliers par blocs pour différents seuils w_s . Nous observons que les deux classes de blocs (bien/mal étiquetés) sont assez faciles à distinguer dans tous les cas.

Il reste finalement à choisir γ . Un γ faible se traduira par des ajouts de couches fréquents. Si une couche a été ajoutée à tort, elle n'aura qu'un effet transitoire car elle devrait ensuite disparaître grâce aux mécanismes exposés plus haut. Cependant, ceci a un coût très important en temps de calcul. Des tests nous ont conduit à opter pour un facteur assez conservateur : $\gamma = 2.5$.

C.2 Choix de la couche erronée.

Dans le cas où nous introduisons une nouvelle couche, il nous faut décider laquelle des deux couches présentes sur chacun des blocs mal étiquetés est concernée.. L'une d'entre elles

FIG. C.2: $\gamma = 1$ ($N_o^l = 153$)

0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1
1	1	1	1	0	0	0	0
1	1	1	1	0	1	0	0
1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1

FIG. C.3: $\gamma = 2$ ($N_o^l = 186$)

0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0
0	1	1	1	0	0	0	0
1	1	0	1	1	1	1	0
1	1	1	1	0	1	1	1

FIG. C.4: $\gamma = 2.5$ ($N_o^l = 202.5$)

0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0
0	1	1	1	0	0	0	0
0	1	0	0	1	1	0	0
0	0	1	0	0	1	0	0

FIG. C.5: $\gamma = 3$ ($N_o^l = 219$)

0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0
0	0	1	1	0	0	0	0
0	1	0	0	1	1	0	0
0	0	1	0	0	1	0	0

FIG. C.6: Blocs détectés comme mal étiquetés (valeur 1) sur la base du nombre d'outliers décomptés sur chaque bloc de la figure C.1, pour différentes valeurs de γ .

98	69	107	94	126	89	102	92
105	143	116	102	119	86	102	109
98	134	114	105	105	99	122	99
74	121	120	92	80	69	102	154
153	177	173	310	106	74	91	126
185	206	222	285	119	155	82	116
194	283	164	192	231	299	191	181
191	195	228	198	176	222	188	197
172	134	198	169	202	160	176	145
181	224	214	187	186	168	177	176
157	205	182	180	180	166	175	152
135	200	193	164	147	141	156	231
218	245	211	397	181	151	167	194
282	297	309	378	210	226	136	181
267	382	242	293	327	389	274	278
278	270	321	273	264	325	266	262

144	102	156	135	170	135	146	124
144	181	177	152	152	131	149	147
131	182	146	148	148	138	147	134
105	165	165	140	122	113	134	199
188	219	198	368	160	121	141	172
251	260	285	338	168	202	116	151
242	345	209	258	285	353	240	242
240	234	287	246	231	276	230	236
202	172	238	202	252	199	216	181
204	266	250	236	209	199	216	201
192	245	230	220	222	196	213	191
173	231	240	211	186	187	189	266
251	287	256	436	214	193	195	245
325	345	357	411	249	269	174	215
304	427	299	328	354	427	302	322
324	308	359	313	310	368	315	297

FIG. C.7: Décompte des outliers de la figure C.1 pour différents seuils w_s . De gauche à droite et de haut en bas : 0, 0.2, 0.4 et 0.6.

a en effet une réalité physique, l'autre pas. Pour les différencier, nous nous appuyons sur le même raisonnement que celui qui nous a permis de calculer des coefficients de confiance pour les déplacements calculés par la mise en correspondance par blocs en 3.4.3.2.

Considérons le résidu développé :

$$\begin{aligned}
r(\mathbf{p}, \mathbf{w}_1, \mathbf{w}_2) &= I(\mathbf{p}, t+1) - I(\mathbf{p} + \mathbf{w}_1, t) \\
&\quad - I(\mathbf{p} + \mathbf{w}_2, t) + I(\mathbf{p} + \mathbf{w}_1 + \mathbf{w}_2, t-1) \quad (\text{C.4}) \\
&= I_1(\mathbf{p}, t+1) - I_1(\mathbf{p} + \mathbf{w}_1, t) \\
&\quad - I_1(\mathbf{p} + \mathbf{w}_2, t) + I_1(\mathbf{p} + \mathbf{w}_1 + \mathbf{w}_2, t-1) \\
&\quad + I_2(\mathbf{p}, t+1) - I_2(\mathbf{p} + \mathbf{w}_1, t) \\
&\quad - I_2(\mathbf{p} + \mathbf{w}_2, t) + I_2(\mathbf{p} + \mathbf{w}_1 + \mathbf{w}_2, t-1) \quad (\text{C.5})
\end{aligned}$$

Supposons par exemple que \mathbf{w}_1 soit l'estimé du mouvement d'une couche présente en \mathbf{p} (et que \mathbf{w}_1 corresponde à la couche I_1), et que \mathbf{w}_2 n'ait aucune réalité physique en \mathbf{p} . Alors, modifier la valeur de \mathbf{w}_1 autour de son estimée changera notablement la valeur du résidu puisque la contribution des quatre premiers termes de (C.5) ne sera plus minimale. Par contre, des variations de \mathbf{w}_2 auront un impact faible puisque de toute façon le deuxième mouvement présent en \mathbf{p} n'était pas correctement appréhendé.

Pour chaque bloc considéré comme mal étiqueté,

- nous calculons la valeur moyenne $r_s^{\mathbf{w}_1}$ du résidu sur le bloc s en faisant varier \mathbf{w}_1 à \mathbf{w}_2 donné¹.
- nous calculons la valeur moyenne $r_s^{\mathbf{w}_2}$ du résidu sur le bloc s en faisant varier \mathbf{w}_2 à \mathbf{w}_1 donné.

Si $r_s^{\mathbf{w}_1}$ est plus grand que $r_s^{\mathbf{w}_2}$, c'est la couche fournissant le mouvement \mathbf{w}_1 qui était présente sur ce bloc. Et réciproquement.

C.3 Estimation du modèle affine de la nouvelle couche.

Nous initialisons l'estimation du modèle affine pour la couche nouvellement introduite en deux temps :

- nous commençons par exploiter les déplacements (issus de la mise en correspondance par blocs) associés dans chaque bloc mal étiqueté à la couche à reconstruire. Nous effectuons ainsi une première estimation robuste du modèle affine complet en suivant l'approche proposée en 3.4.3.1.
- nous utilisons ensuite la méthode d'estimation paramétrique *restreinte à la nouvelle couche seulement*. Reprenons (4.3) :

$$J(\Theta) = \sum_{\mathbf{p} \in \mathfrak{S}} r(\mathbf{p}, \theta_{e_{\mathbf{p}}(1)}, \theta_{e_{\mathbf{p}}(2)})^2 \quad (\text{C.6})$$

$$\text{avec } \Theta = (\theta_1, \theta_2, \dots, \theta_n, \theta_{n+1}) \quad (\text{C.7})$$

Quand les étiquetages et les paramètres de mouvement $\theta_1, \theta_2, \dots, \theta_n$ des n couches existantes sont fixés, (C.6) devient une fonction ne dépendant que du nouveau modèle introduit θ_{n+1} . Nous utilisons alors la technique incrémentale de Gauss-Newton pour estimer θ_{n+1} . Partant de l'initialisation précisée plus haut, nous utilisons une pyramide (B-spline) des images, et affinons les estimés en minimisant itérativement une version de (C.6) linéarisée au second ordre.

¹En pratique, nous faisons varier chaque composante de \mathbf{w}_1 de -4 et $+4$ pixels. L'utilisation de déplacements entiers pour le calcul de la moyenne évite d'avoir à interpoler de nouveau les images.

Une fois le modèle de mouvement associé à la nouvelle couche estimé, nous itérons la minimisation de la fonction d'énergie (5.1) avec cette couche additionnelle. Ceci nous permet :

- en cas d'ajout pertinent de couche, de lui attribuer l'ensemble des blocs qui lui correspondent (il est fort probable que certains n'aient pas encore été repérés) et d'améliorer encore la qualité l'estimation des paramètres de mouvement.
- en cas d'ajout de couche intempestif, de la faire disparaître ou bien de la fusionner à une autre couche.

C.4 Utilisation pratique de l'ajout de couches

L'ajout d'une couche mène donc quasiment à un doublement du temps de calcul de la méthode d'estimation de mouvement dans son ensemble. Nous cherchons donc à rendre son utilisation aussi peu fréquente que possible.

Ainsi n'effectuons nous ce test qu'une fois seulement, à la fin de l'exécution normale du schéma markovien. Ceci permet de se baser sur une segmentation et des modèles de mouvement fiables, qui limiteront le risque d'ajouter une couche à mauvais escient.

D'autre part, nous faisons en sorte d'être le plus souvent possible en situation de sur-segmentation lors de l'initialisation, et le moins souvent possible en situation de sous-segmentation. Outre le fait que le premier cas est infiniment plus facile à traiter que le second, les conséquences pratiques d'une sur-segmentation pour l'application finale sont souvent moins graves puisqu'elle n'empêche pas d'estimer correctement les mouvements.

Ceci explique les réglages choisis de la transformée de Hough², qui mènent à de fréquentes sur-segmentations, mais permettent de n'oublier que très rarement des couches vraiment présentes dans l'image.

Nous avons implanté l'approche proposée, et nous l'avons testé sur 100 générations de séquences d'images en situation de bitransparence distribuée (comme expliqué en 4.3.1.1). Nous présentons les précisions des modèles de mouvements estimés, mesurés avec les trois métrique de 5.5.1.1 sur le tableau C.2.

TAB. C.2: Précision de l'estimation jointe de mouvement et de segmentation en situation de bi-transparence distribuée, pour différent réglage du seuil détectant les blocs suspects lors de l'ajout de couche. Pour comparaison, le RF calculé avec les vitesses réelles vaut 6,22 sur les images diagnostiques, et 6,37 sur les images interventionnelles.

γ	Diagnosticque			Interventionnel			Temps de Calcul
	EQM	EEM	RF	EQM	EEM	RF	
1	0,76	0,53	6,82	2,27	1,44	9,35	325
2	0,78	0,52	6,84	2,38	1,50	9,77	291
3	0,76	0,51	6,84	2,23	1,44	9,61	195
5	0,75	0,51	6,87	2,29	1,45	9,58	164

Nous constatons donc qu'avec ce type de réglages, le mécanisme d'ajout de couches

²Pas de discrétisation relativement grossier, et extraction de paramètres considéré révélateur même quand seulement 5 blocs sont concernés.

n'améliore quasiment pas la précision d'estimation. Les résultats rapportés au tableau C.2 mettent en évidence un gain infime de précision quand γ est faible (et donc que des couches sont "faciles" à ajouter) au prix d'une augmentation conséquente du temps de calcul. Quand γ est plus élevé, l'augmentation des coûts de calcul est en moyenne plus raisonnable, mais la précision d'estimation est identique à une version sans ajout de couches.

Voilà pourquoi nous avons choisi de ne pas intégrer ce mécanisme d'ajout de couche dans la version finale de l'algorithme d'estimation de mouvements en situation de bi-transparence distribuée. Mais le traitement d'autres types d'images pourrait mener à d'autres réglages de l'algorithme, qui profiteraient de la méthode d'ajout de couches.

Chapitre D

Détection des configurations mono- et multi-couches

Nous avons évoqué en section 5.4 la possibilité de détecter les blocs en configuration mono- ou multi-couches, et d'intégrer cette information au sein de notre formalisme d'estimation de mouvement en situation de bi-transparence distribuée. Nous avons aussi mentionné que ces méthodes, efficaces sur des couches texturées, ne sont pas exploitables sur les images cliniques peu texturées que nous considérons.

Nous présentons de manière détaillée dans cette annexe les méthodes de détection de configuration que nous avons développées, et expliquons comment nous pouvons modifier notre schéma d'estimation en situation de bi-transparence distribuée pour tenir compte de cette information. Nous revenons sur les limitations de cette approche quand il s'agit de traiter des images cliniques.

D.1 Traitement des blocs en configuration mono-couche

Nous procédons en deux temps : nous commençons par repérer les blocs susceptibles d'être en situation mono-couche, avant d'intégrer cette information dans le formalisme markovien par l'intermédiaire d'une fonction $\eta(\cdot)$.

D.1.1 Détection des blocs suspects

Pour repérer les blocs mono-couches, nous cherchons à détecter les vitesses estimées ne correspondant à aucune réalité physique. L'idée est la suivante : si la valeur du résidu sur le bloc s

$$\nu(\theta_{e_i(s)}, \theta_{e_j(s)}, s) = \sum_{\mathbf{p} \in s} r(\mathbf{p}, \theta_{e_i(s)}, \theta_{e_j(s)}) \quad (\text{D.1})$$

varie peu pour différentes valeurs de $\theta_{e_j(s)}$ (en gardant $\theta_{e_i(s)}$ constant), la couche correspondant à $e_j(s)$ est peu représentative. Il n'y a alors probablement qu'une couche sur s , correspondant à l'étiquette $e_i(s)$.

Plus précisément, pour détecter une éventuelle situation de mono-couche correspondant à $e_i(s)$, nous calculons la valeur moyenne $\bar{\nu}(\theta_{e_i(s)}, s)$ du résidu $\nu(\theta_{e_i(s)}, \cdot, s)$ en appliquant n

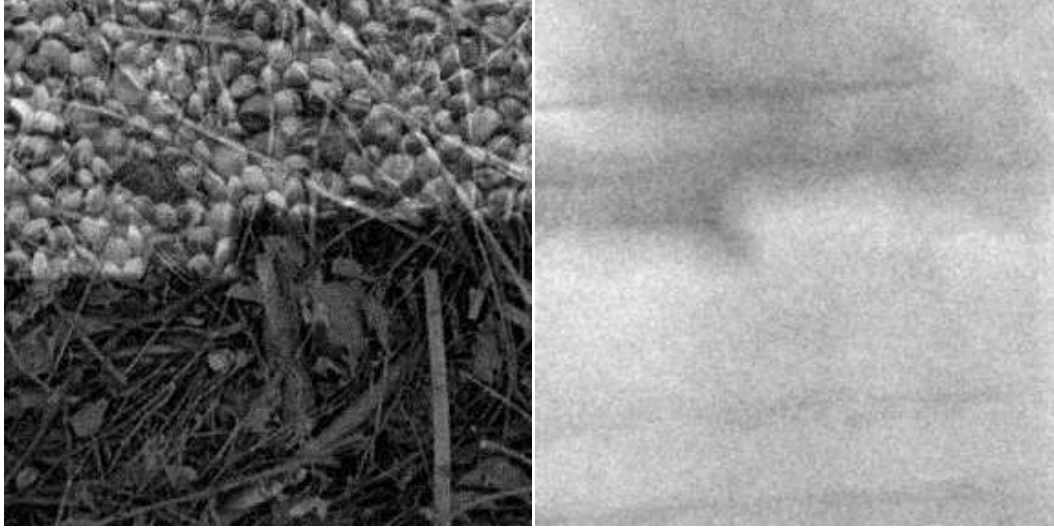


FIG. D.1: Image en situation de bi-transparence (en haut) et mono-transparence (en bas). À gauche, les couches sont formées par des images de textures, à droite par des images cliniques réalistes. La segmentation est la même, et les vitesses des couches sont identiques.

mouvements à la seconde couche (définis par θ_k^{test} , $k = 1, \dots, n$)¹.

$$\bar{\nu}(\theta_{e_i(s)}, s) = \frac{1}{n} \sum_{k=1}^n \nu(\theta_{e_i(s)}, \theta_k^{test}) \quad (\text{D.2})$$

Les tableaux D.1 et D.2 contiennent les valeurs que prend cet indicateur sur un exemple de séquence (textures en translation, avec ajout de bruit, séquence de gauche de la Fig.D.1). La séquence est en configuration bi-couche sur sa partie supérieure, et mono-couche sur la partie inférieure. Le fait qu'un des deux mouvements soit sans signification physique apparaît clairement sur la partie inférieure de l'image. Nous reviendrons en D.3 sur le choix d'images de textures (et non cliniques) pour illustrer notre méthode.

Pour conclure sur la présence ou non d'une situation mono-couche, il nous faut décider si $\bar{\nu}(\theta_i, s)$ est significativement différent du résidu minimal $\nu(\theta_{e_i^*(s)}, \theta_{e_j^*(s)}, s)$, où $(e_i^*(s), e_j^*(s))$ désigne l'étiquetage bi-couche donnant le résidu minimum sur le bloc.

Nous souhaitons éviter un seuillage brutal autour d'une valeur fixe, car le seuil doit être différent selon que l'on traite des zones texturées ou non. Il est spécifié en ligne comme suit. Nous calculons d'abord la valeur médiane de $\nu(\theta_{e_i^*(s')}, \theta_{e_j^*(s')}, s')$ sur un voisinage de s :

$$\nu_{med}(s) = \underset{s' \in V(s)}{med} \nu(\theta_{e_i^*(s')}, \theta_{e_j^*(s')}, s') \quad (\text{D.3})$$

puis, son écart médian

$$\Delta \nu_{med}(s) = \underset{s' \in V(s)}{med} |\nu(\theta_{e_i^*(s')}, \theta_{e_j^*(s')}, s') - \nu_{med}(s)|. \quad (\text{D.4})$$

¹En pratique, nous les construisons en ajoutant ou soustrayant 4 pixels aux composantes translationnelles de $\theta_{e_j(s)}$ (ce qui évite d'avoir à calculer de nouvelles interpolations).

TAB. D.1: Valeurs de $|\bar{\nu}(\theta_{e_1(s)}, s) - \nu(\theta_{e_1^*(s)}, \theta_{e_2^*(s)}, s)|$ sur les blocs de la séquence de gauche de la fig.D.1. Les deux modèles estimés correspondent aux deux mouvements présents. Il apparaît clairement que la première étiquette ne correspond à aucune réalité physique sur la partie inférieure de l'image.

13310	11309	12492	11198	14606	13051	13739	14058
14535	12547	13153	11446	10826	10179	12828	13373
14636	10044	14426	13946	13427	13984	9579	12394
13589	10826	11388	5906	3405	1575	1270	2311
1612	4063	1327	1506	-64	-57	46	-54
-43	-5	9	-8	8	9	-48	54
-60	115	-32	47	-106	68	23	-29
-40	-38	4	6	-20	-34	-12	17

TAB. D.2: Valeur de $|\bar{\nu}(e_2(s), s) - \nu(\theta_{e_1^*(s)}, \theta_{e_2^*(s)}, s)|$ sur les blocs de la séquence de gauche de la fig.D.1. Les deux modèles estimés correspondent aux deux mouvements présents. La vitesse testée correspond à une réalité physique sur tous les blocs.

9763	7615	11576	8798	7598	7636	6937	10205
10559	6175	8230	8841	7303	6088	7972	9175
5994	4442	5923	5888	9396	9590	8370	10367
7675	3364	5686	9526	7812	8383	6914	8692
6912	5708	5329	11003	6452	6465	13854	13594
7967	6827	7671	9834	6528	7788	12368	6951
5949	7016	6608	5845	5698	5025	9227	8380
6192	4258	7230	9472	9369	5573	10191	9729

La taille du voisinage doit être choisie assez petite pour que les indicateurs soient typiques des textures du bloc s , mais assez grande pour qu'une zone mono-couche éventuellement présente autour de s ne fausse pas les statistiques. En pratique, nous choisissons $7 \times 7 (= 49)$ blocs.

$\Delta\nu_{med}(s)$ représente donc la variation typique des résidus minimaux autour de la valeur représentative $\nu_{med}(s)$ de ce minimum. Elle nous permet de fixer l'ordre de grandeur des variations du résidu qui ne sont pas significatives dans ce voisinage.

Le mécanisme de décision est le suivant. si

$$\left| \bar{\nu}(\theta_{e_i(s)}, s) - \nu(\theta_{e_i^*(s)}, \theta_{e_j^*(s)}, s) \right| < \alpha \Delta\nu_{med}(s), \quad (\text{D.5})$$

nous considérons que la couche correspondant à $e_j(s)$ n'est probablement pas significative. En d'autres termes, si, en changeant la vitesse associée à l'étiquette $e_j(s)$, l'ordre de grandeur du résidu reste le même, nous considérons le bloc comme potentiellement mono-couche.

D.1.2 Intégration de la détection mono-couche

Nous parlons de situation mono-couche *potentielle*, car nous préférons ne pas trancher sur cette question indépendamment des autres critères guidant la segmentation aux couches. Il est en particulier crucial de ne pas découpler la détection mono-couche des contraintes de voisinage auxquelles elle est intimement liée.

Voilà pourquoi nous intégrons le critère de détection mono-couche dans le formalisme markovien en ajoutant une fonction $\eta(\cdot)$ dans l'expression (5.1)

$$\begin{aligned} F(e, \Theta) &= \sum_{s \in S} \left(\sum_{\mathbf{p} \in s} \rho(r(\mathbf{p}, \theta_{e_1(s)}, \theta_{e_2(s)})) - \mu \cdot \eta(s, e_1(s), e_2(s)) \right) \\ &+ \mu \sum_{\langle s, t \rangle \in C} \left((1 - \delta(e_1(s), e_1(t))) (1 - \delta(e_1(s), e_2(t))) \right. \\ &\quad \left. + (1 - \delta(e_2(s), e_1(t))) (1 - \delta(e_2(s), e_2(t))) \right) \end{aligned} \quad (\text{D.6})$$

La minimisation de la fonction globale $F(e, \Theta)$ sans addition de ce terme supplémentaire conduit sur un bloc mono-couche à l'estimation correcte du mouvement présent sur le bloc, mais aussi à l'ajout d'un modèle de mouvement fictif. Ce dernier est associé à une étiquette existante, déterminée donc "au hasard" selon le contexte et les informations disponibles à ce stade.

Or, nous voudrions qu'un bloc mono-couche s correspondant à $e_i(s)$ soit reconnu comme tel, par exemple en favorisant l'étiquetage $(e_i(s), e_i(s))$. Comme le schéma d'estimation de la bi-transparence distribuée ne favorisera aucunement cet étiquetage vis à vis des $(e_i(s), e_j(s))$ ou $(e_j(s), e_i(s))$ ($i \neq j$), nous choisissons d'ajouter dans (5.1) un terme, $-\mu \cdot \eta(s, e_i(s), e_j(s))$

- valant 0 lorsque la configuration testée est réellement bi-couche $(e_i(s), e_j(s))$ avec $i \neq j$;
- valant 0 lorsque la configuration testée est mono-couche "incertaine", du type $(e_i(s), e_i(s))$, mais ne vérifiant pas la condition (D.5) ;
- valant $-\mu$ lorsque la configuration testée est mono-couche "probable", du type $(e_i(s), e_i(s))$ et vérifiant le (D.5).

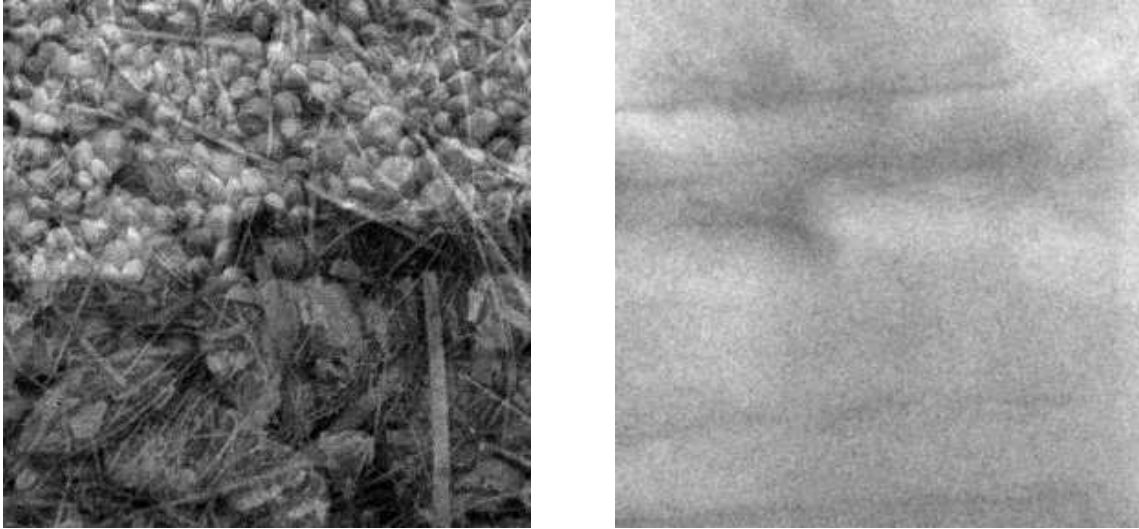


FIG. D.2: Images en situation de tri-transparence (en haut) et bi-transparence (en bas). À gauche, les couches sont formées par des images de textures, à droite par des images cliniques réalistes. La partition est la même, et les mouvements des couches sont identiques.

En pratique, ceci ne change rien à la phase d'estimation des modèles de mouvement à étiquetage fixé. Par contre, lorsque les modèles de mouvement sont fixés et que l'on teste tous les étiquetages possibles pour un bloc par la technique ICM, le terme ajouté permet de favoriser l'hypothèse mono-couche ($e_i(s), e_i(s)$) si le test (D.5) correspondant est vérifié.

D.2 Détection des blocs en configuration de multi-couches

De la même façon, nous procédons pour la détection des blocs en configuration multi-couches en deux étapes : la détection de blocs suspects, puis l'utilisation effective de cette information dans l'algorithme.

D.2.1 Détection des blocs suspects

En situation de multi-couches, le résidu n'a pas de sens physique. Le minimum atteint ne recouvre pas de réalité et est donc peu significatif. L'idée est cette fois-ci d'évaluer la significativité du minimum vis à vis des valeurs typiques de résidus obtenus sur l'image. Nous construisons le critère :

$$\nu(\theta_{e_i^*(s)}, \theta_{e_j^*(s)}, s) > \nu_{med}(s) + \alpha' \Delta \nu_{med}(s). \quad (D.7)$$

S'il est vérifié, nous considérons que le minimum n'est pas significatif, et qu'aucun des deux mouvements estimés n'est fiable. Tab.D.3 contient les valeurs des résidus sur une séquence d'images de textures en situation de bi-transparence et de tri-transparence. Les résidus correspondant à la situation de tri-couches sont clairement non représentatifs.

Nous reprenons les grandeurs précédentes $\nu_{med}(s)$ et $\Delta \nu_{med}(s)$ par simplicité, mais nous pouvons éventuellement les calculer sur des voisinages plus petits (car la situation multi-couche est plus rare et plus localisée que la situation mono-couche). Notons qu'en outre

TAB. D.3: Valeurs du résidu sur les blocs de la séquence de textures représentée sur l'image de gauche de la fig.D.2. Les deux modèles estimés correspondent aux deux mouvements présents sur toute l'image. Nous observons un résidu non significatif sur les blocs supérieurs de l'image.

14506	13655	16477	12162	15876	16261	16125	19402
18453	13710	18627	14918	19999	16346	14367	17389
16274	16363	20524	16750	15659	17516	15409	20019
24629	18936	15556	12593	8881	7476	6972	9166
10000	12210	8517	4303	1556	1640	1588	1623
1609	1599	1423	1600	1626	1695	1541	1555
1498	1641	1659	1519	1492	1588	1510	1732
1581	1593	1552	1526	1537	1489	1608	1574

(D.7) permet de détecter aussi les blocs bi-couches sur lesquels l'estimation du mouvement est mauvaise.

D.2.2 Intégration de la détection de configurations de multi-couches

Alors que le schéma d'estimation en bi-transparence distribuée permet d'appréhender les situations mono-couches, il n'est par contre pas envisageable de l'étendre facilement aux situations multi-couches.

Quand un bloc est détecté en situation de multi-couches, nous cherchons à empêcher ses interactions avec ses voisins, pour éviter de transmettre des informations peu fiables par le biais du lissage spatial. Ainsi, lors de la minimisation par ICM à modèles de mouvement fixés, les cliques comprenant un bloc bi-couche exclueront les voisins en multi-couches.

Par contre, les blocs multi-couches eux-mêmes seront traités normalement. L'étiquetage obtenu peut en effet être exploitable, que ce soit parce qu'une des trois couches est en fait homogène et donc a permis l'estimation du mouvement des deux autres, ou parce que les contraintes de lissage ont imposé des étiquettes cohérentes spatialement. Ce sera à l'utilisateur de choisir s'il préfère une carte de segmentation finale mettant en évidence les zones multi-couches, ou une segmentation présentant les étiquettes mono et bi-couches les plus probables.

Ce choix est d'ailleurs sans conséquences sur la qualité des estimations de mouvement. Grâce à l'utilisation d'estimateurs robustes, quelques blocs multi-couches mal étiquetés ne seront pas gênants lors de la minimisation paramétrique. Nous avons d'autre part interdit aux blocs suspects de produire des contraintes de lissage sur leurs voisins.

D.3 Discussion sur la pertinence de la détection de configurations

Nous résumons dans le Tab.D.4 les comportements de l'algorithme sur les blocs mono- et multi-couches, les méthodes que nous proposons pour détecter ces configurations, ainsi que les répercussions dans l'algorithme joint d'estimation et de segmentation des mouvements

transparentes dans son ensemble.

Comme évoqué en 5.4.3, les stratégies de prise en compte et de détection que nous avons

TAB. D.4: Résumé des procédures de détection des situations mono-couches et multi-couches

Mono-couche	Comportement du résidu	Minimum quand le mouvement présent est bien estimé, quelle que soit la valeur de l'autre
	Détection	$\left \frac{1}{n} \sum_{k=1}^n \nu(\theta_{e_i(s)}, \theta_k, s) - \nu(\theta_{e_i^*(s)}, \theta_{e_j^*(s)}, s) \right < \alpha \Delta \nu_{med}(s)$
	Intégration	Bonus de $-\mu$ par l'intermédiaire de la fonction $\eta(e_i(s), e_j(s))$, quand $i = j$ et que le test de détection est vérifié relativement à $e_i(s)$. Ceci pour favoriser l'hypothèse d'étiquetage mono-couche sur ce bloc.
Multi-couches	Comportement du résidu	Incohérent. Ignoré grâce au formalisme robuste. Étiquetage aléatoire sauf si une couche est moins texturée que les autres.
	Détection	$\nu(\theta_{e_i^*(s)}, \theta_{e_j^*(s)}, s) > \nu_{med}(s) + \alpha' \Delta \nu_{med}(s)$
	Intégration	Pas de lissage spatial incluant des blocs multi-couches. Blocs multi-couches eux-mêmes traités normalement.

décrites ne fonctionnent pas de manière fiable sur tous les types d'images.

La détection de la configuration mono-couche exploite l'idée que le comportement du résidu lorsque l'on modifie une des vitesses estimées diffère nettement selon qu'il s'agit d'une vitesse fictive ou réelle.

- *Si la vitesse estimée est fictive*, le résidu ne change pratiquement pas si l'image est peu bruitée. Dans le cas contraire, il ne fluctue que faiblement du fait du bruit de la séquence d'images.
- *Si la vitesse estimée a un sens physique*, le résidu grandit si on la modifie puisqu'on cesse de mettre correctement en correspondance les éléments de la couche réelle. Cependant, lorsque cette couche est peu texturée, les variations du résidu ont une amplitude faible.

Ainsi, lorsque l'on considère des images bruitées et peu texturées, les deux comportements que l'on cherche à différencier ont tendance à se rapprocher. Nous présentons dans Tab.D.5 l'équivalent du contenu de Tab.D.1, à la seule différence que les images utilisées ne sont plus des textures, mais des images cliniques (image de droite de la Fig.D.1). La configuration est la même : la partie supérieure de l'image est en situation de bi-couche, la partie inférieure en mono-couche. Les mouvements et les niveaux de bruit introduits sont les mêmes.

La différence de comportement du résidu moyenné n'est plus révélateur sur la plupart des blocs. Les réalisations du bruit prennent le pas sur l'information faible portée par l'image. Ce phénomène est d'ailleurs encore plus marqué quand les mouvements des couches ne sont qu'imparfaitement estimés.

Nous présentons le même genre de tests pour le critère multi-couche. Le tableau D.6 contient les valeurs des résidus minimaux sur chaque bloc d'une séquence d'images illustrée

TAB. D.5: Valeurs de $\nu(e_1(s), s) - \nu(\theta_{e_i^*(s)}, \theta_{e_j^*(s)}, s)$ sur les blocs d'une séquence formée à partir d'images *cliniques* en situation bi-couche en haut de l'image et mono-couche en bas. Les deux mouvements estimés correspondent aux deux mouvements présents.

-102	6	1	11	-28	156	38	119
91	101	6	2	182	-31	-16	68
56	-25	-15	-41	78	5	-1	-11
0	123	-99	120	165	-64	-40	5
92	136	-97	48	116	-13	66	-121
52	41	106	86	142	44	64	82
-45	66	-42	-48	-7	68	-3	3
37	-23	51	-89	77	22	20	-32

TAB. D.6: Valeurs du résidu sur les blocs d'une séquence formée à partir d'images cliniques comprenant une situation de tri-transparence (en haut) et de bi-transparence (en bas).

2037	1779	1764	1702	1695	1925	1843	1751
2197	1989	1667	2162	1781	2109	2161	1974
2263	2073	1985	2452	2112	2514	2549	2200
2685	2734	2158	1554	1654	1725	1980	1666
1941	2197	1720	1775	1646	1708	1664	1846
1563	1531	1633	1364	1950	1732	1602	1496
1579	1544	1424	1487	1730	1644	1376	1523
1791	1637	1582	1516	1574	1587	1481	1596

par l'image de droite de Fig.D.2, où on a substitué les images de textures par des couches cliniques. La partie supérieure de l'image est en configuration tri-couche, et la partie inférieure en situation bi-couche.

L'ordre de grandeur du résidu (qui est divisé ici par la taille du bloc) est logique : le bruit entachant chaque image a un écart-type de 20, et donc le résidu, homogène à une variance, est de l'ordre de $4 \times 20^2 = 1600$ puisqu'il est construit comme somme de quatre images.

Les zones tri-couches sont repérables, mais la différence est bien moins nette que dans Tab.D.3 où il avait un rapport 5 entre les deux cas. Le seuillage est donc délicat, et devient même impossible quand les mouvements ne sont pas parfaitement connus mais estimés avec une précision de l'ordre du pixel.

Plus intuitivement, la figure 5.4 illustre la difficulté de la tâche de détection de ces configurations. Elle montre une image clinique sur laquelle trois blocs sont dessinés. L'un d'entre eux (le bleu) est en situation mono-couche, correspondant aux poumons. Le bloc vert est en situation bi-couche (poumons+côte) avec deux couches relativement texturées. Sa configuration exacte pourrait être détectée. Par contre, le bloc rouge se trouve lui-aussi en situation bi-couche, mais la couche du cœur y est si homogène qu'elle ne sera pas détectable indépendamment des autres.

Bien sûr, le lissage spatial permet de diffuser l'information d'étiquetage et donc de pallier en partie ce problème. C'est ainsi que nous sommes en mesure de présenter des résultats

convaincants de détection mono-couche sur des images vidéo dans l'annexe E. Cependant, les images cliniques contiennent trop peu de blocs peu texturés pour que les détections de ces configurations puissent y être effectuées de manière toujours fiable.

Nous choisissons donc d'y renoncer, et ce d'autant que la version de l'algorithme original est robuste aux situations mono-couches et bi-couches :

- les mouvementss des blocs mono-couches sont bien estimés,
- les blocs multi-couches ne gênent pas directement les estimations des modèles de mouvement.

Comme l'application que nous visons (le débruitage compensé en mouvement) nous autorise à accepter une carte de segmentation imparfaite pourvu que les mouvements vraiment présents soient bien estimés, nous préférons utiliser l'algorithme sans adaptation quand il s'agit de traiter les images médicales.

Chapitre E

Traitement d'images vidéos transparentes

Le phénomène de transparence est également présent dans des séquences vidéos classiques. Nous avons illustré quelques situations de ce type à la Fig.2.1. Il peut s'agir de réflexions dans un miroir ou sur une vitre, d'observations derrière un objet translucide, ou d'effets spéciaux (effet "fantôme", fondu-enchaîné, etc).

Nous présentons dans cette annexe quelques traitements de séquences de ce type, dont certains résultats ont été publiés dans [Auvray 06a].

E.1 Une séquence de bi-transparence

Cette première séquence représente un visage se reflétant dans un cadre en mouvement (Fig.E.1)¹. On observe la superposition de ce visage et de la femme prise en photo. Les mouvements imprimés manuellement sont souples mais irréguliers : le cadre subit une rotation. En particulier, la constance des mouvements sur deux intervalles de temps successifs n'est pas vérifiée.

Nous rassemblons les résultats de notre méthode d'estimation de mouvements transparents à la figure E.2. Bien que la séquence (redimensionnée pour que le bord du cadre ne soit pas visible) soit en situation de bi-transparence simple, nous avons utilisé notre algorithme dédié à la bi-transparence distribuée (adapté à la mono- et multi-transparence - Annexe D).

Nous montrons la segmentation spatiale des images en couches, les champs de mouvement estimés, et les images de différence recalées par rapport aux deux champs estimés. Ces dernières permettent de juger de la qualité d'estimation du mouvement : si elle est correcte, la couche correspondante doit disparaître et on ne doit plus observer que le "gradient" de la deuxième.

Nous observons que la segmentation en deux couches est satisfaisante. À la date 2, deux blocs très peu texturés ont été considérés en situation de mono-couche, ce qui est acceptable étant donné l'absence de reflets observés. Les images de différences montrent que les mouvements estimés sont corrects.

La segmentation en couches et l'estimation des mouvements transparents sont adéquates à

¹La séquence, due à Mickael Black, se trouve sous <http://www.cs.brown.edu/people/black>



FIG. E.1: Images 0, 2, 4, 6, et 8 de la séquence en situation de bi-transparence, qui compte 10 images.

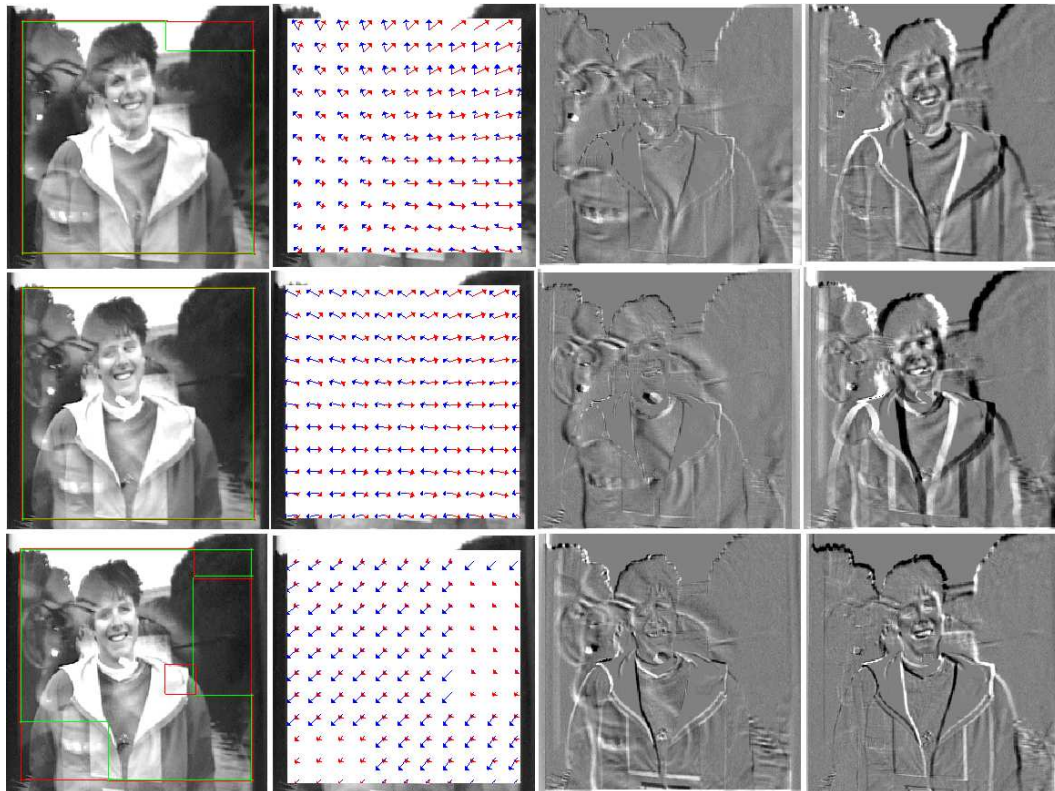


FIG. E.2: Résultats de l'estimation des mouvements transparents aux dates 2 et 4. De gauche à droite : segmentation de l'image en couches (délimités par les lignes de couleurs), champs estimés (amplitude des vecteurs tracés multipliée par 2 par rapport au déplacements calculés), images des différences recalées pour chaque modèle de mouvement estimé (voir texte principal).



FIG. E.3: Trois images successives d'une séquence synthétique en bi-transparence distribuée.

la date 4. Les problèmes dus à la détection des situation de mono-couche sont plus marqués à la date 6 : deux groupes de blocs ne sont attribués qu'à la photo et non au visage. Si c'est très compréhensible sur le groupe de droite, sur lequel la couche du visage est quasiment uniforme, c'est plus étonnant sur le groupe de gauche. Le problème vient du fait que le mouvement du cadre change brusquement sur ces trois images, ce qui perturbe évidemment l'algorithme. Cela dit, la configuration de l'image est correctement approchée, et le mouvement de la couche correspondant au visage est bien estimé.

E.2 Une séquence synthétique de bi-transparence distribuée

Nous décrivons dans cette section le traitement d'une séquence synthétique en situation de bi-transparence distribuée (Fig.E.3). Deux portraits y sont ajoutés à un panorama en translation verticale. L'un d'entre eux (Léna) subit un zoom ; l'autre (Barbara) une translation horizontale. Cette séquence a été construite pour traiter un exemple typique de bi-transparence distribuée à vérité-terrain connue. Il nous permet également d'étudier comment se déroule la détection de situation de mono-couche.

Les résultats sont contenus dans la Fig.E.4. Nous y présentons la segmentation en couches, ainsi que la segmentation en régions bi-couches (c'est-à-dire celles qui correspondent à des couples d'étiquettes différents). Nous montrons également les mouvements estimés, et les images de différence recalées.

Ces dernières montrent que les estimations des mouvements des trois couches sont excellentes. En effet l'erreur d'estimation est de 0.11 pixels seulement. La localisation des blocs bi-couches est intéressante mais imparfaite. Même sur un exemple synthétique contrôlé, il est difficile de détecter exactement la configuration de chaque bloc. La couche correspondant au panorama contient notamment des zones aux textures très différentes (du ciel quasiment homogène aux rochers très contrastés), ce qui rend le calcul du seuil de détection des configurations monocouches peu robuste. C'est ainsi qu'une fausse situation de mono-couche correspondant à Léna est détectée sur le ciel.

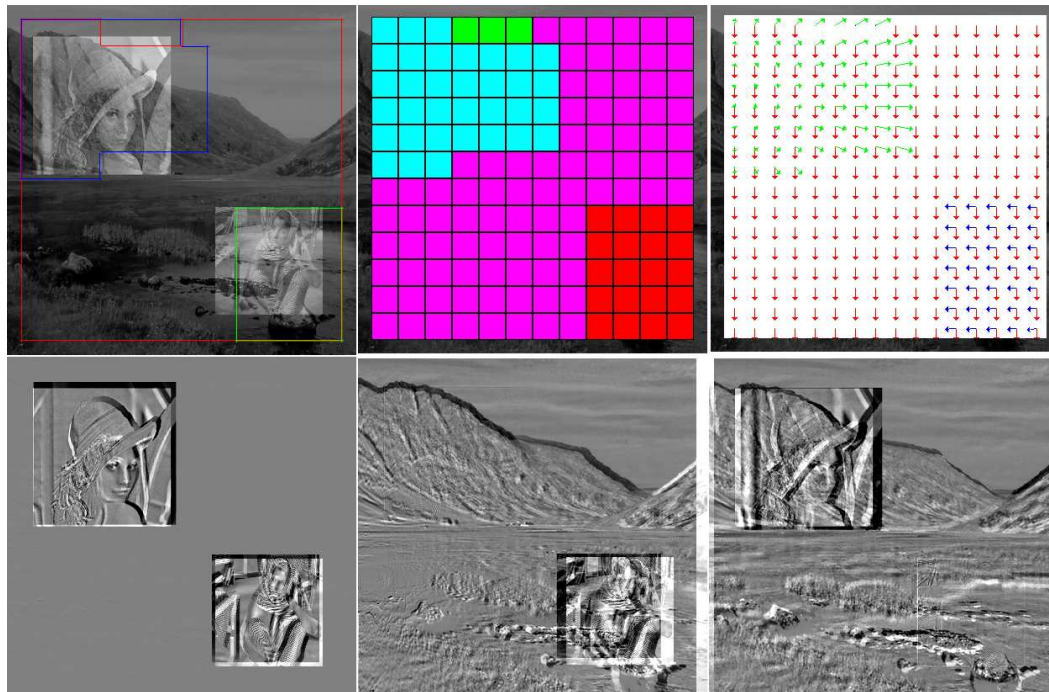


FIG. E.4: Résultats de l'estimation des mouvements transparents et de la segmentation en couches sur un exemple synthétique de bi-transparence distribuée construit à partir d'images vidéo. En haut à gauche : la segmentation en couches, au centre la segmentation en régions bi-couches, à droite les mouvements estimés (amplitude des vecteurs tracés multipliée par 2 par rapport aux déplacements calculés). En bas, les images de différences recalées respectivement par rapport aux modèles de mouvements associés au paysage, à Léna et à Barbara. On peut constater que la couche correspondant au mouvement compensé est parfaitement éliminée dans l'image de différences recalées.



FIG. E.5: Images aux dates 0, 10 et 20 de la séquence de 30 images Mona Lisa.

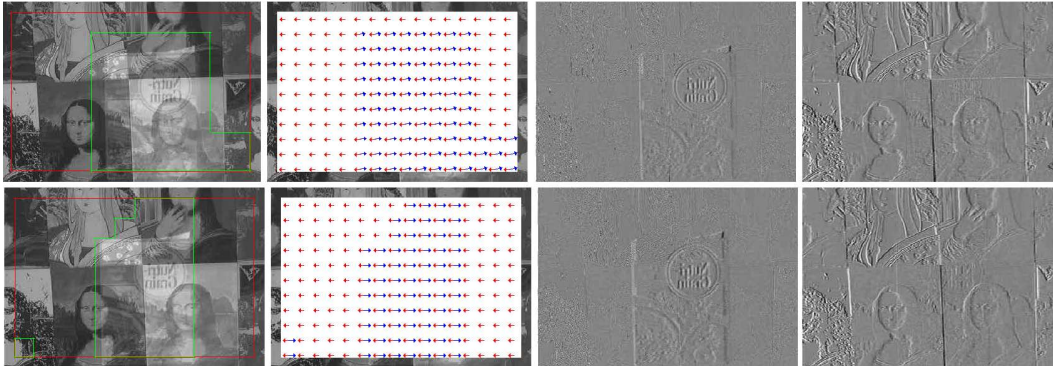


FIG. E.6: Résultats de l'estimation et de la segmentation aux dates 10 et 20. De gauche à droite : la segmentation en couches, les champs issus des modèles de mouvement estimés (amplitude des vecteurs multipliés par 2 par rapport aux déplacements calculés), et les images de différence recalées.

E.3 Une séquence de laboratoire de bi-transparence distribuée

Nous avons ensuite considéré une séquence réelle de laboratoire en situation de bitransparence distribuée (Fig.E.5)². Elle représente une boîte de corn-flakes se reflétant dans la vitre protégeant un tableau de patchworks de Mona-Lisa. La caméra suit un mouvement de translation relativement régulier, qui se traduit par deux mouvements apparents de translation opposés dans l'image.

Nous présentons les résultats de notre algorithme à la Fig.E.6 : la segmentation en couches, les champs issus des modèles de mouvement estimés et les images de différences recalées. Là aussi, les mouvements sont très bien estimés. La segmentation est exploitable mais grossière du fait de la difficulté à détecter les configurations mono-couches précisément.

E.4 Une séquence réelle de bi-transparence distribuée

Nous traitons pour finir une séquence de transparence issue d'un film (Fig.E.7). Il s'agit à notre connaissance du premier exemple de traitement d'une séquence de transparence

²La séquence, due à Yu, se trouve sous <http://www.ks.informatik.uni-kiel.de/wy/motiondemo.html>

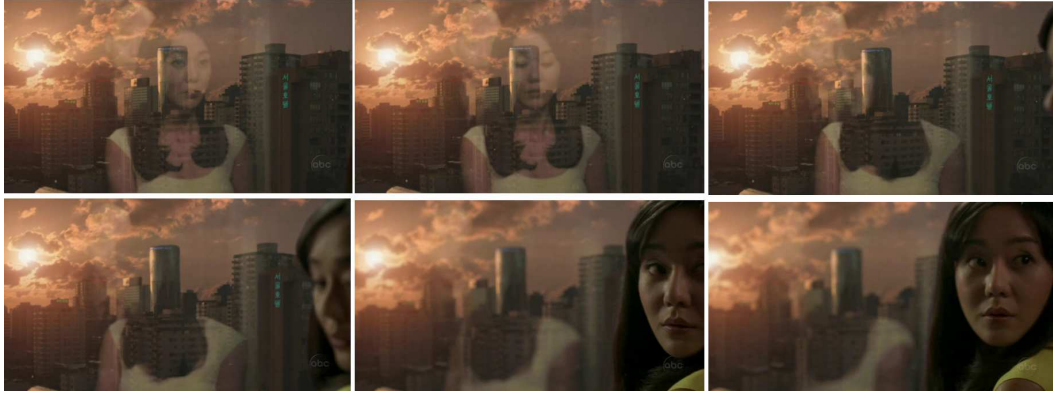


FIG. E.7: Images aux dates 0, 20, 40, 50, 60 et 70 de la séquence de 80 images Lost.

réelle qui ne soit pas acquise en laboratoire à cette fin.

Elle contient deux personnages se reflétant sur la vitre de la fenêtre de leur appartement, qui donne sur la ville. La caméra est dans un premier temps statique, puis effectue un mouvement souple de translation et de rotation pour faire entrer le visage de l'héroïne au premier plan. Ce mouvement de caméra fait apparaître le panorama de la ville et les personnages en translation.

Nous montrons les résultats obtenus à la date 55 à la Fig.E.8. Les résultats sont moins bons à d'autres dates, du fait notamment de mouvements parasites des personnages et du visage apparaissant au premier plan.

Sur l'exemple présenté, le premier plan occupe toutefois suffisamment peu de blocs pour être ignoré. Si ce n'avait pas été le cas et que son mouvement pouvait être représenté par un modèle paramétrique, sa présence n'aurait d'ailleurs pas beaucoup gêné l'algorithme, qui y aurait vu une zone monocouche. Seule la zone ainsi occultée n'aurait pas respecté la modélisation de l'image, mais le formalisme robuste aurait suffi à ignorer ces outliers. Cependant dans le cas qui nous occupe, l'héroïne tourne la tête rapidement, ce qui se traduit par un mouvement impliquant de nombreuses occultations.

Sur la séquence traitée, les mouvements sont bien estimés. Dans la première image de différence recalée le reflet des personnages (même s'ils restent légèrement visibles du fait du mouvement de rotation de l'héroïne), et dans la seconde l'image de la ville a bien disparu. La segmentation est également pertinente, à l'exception de deux blocs homogènes du premier plan qui sont attribués à la couche de reflet des personnages.

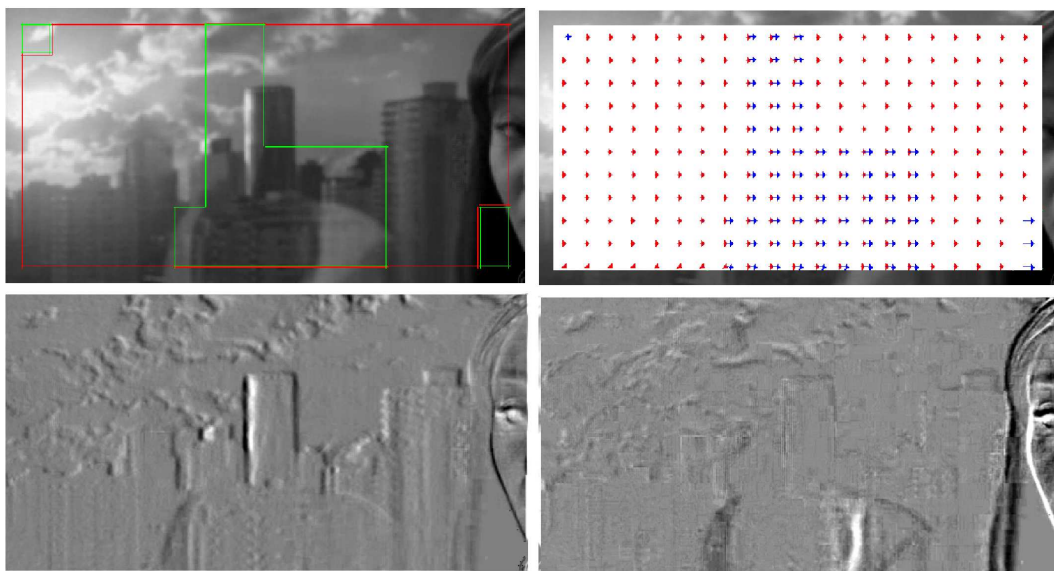


FIG. E.8: Résultats de l'estimation et de la segmentation jointe à la date 55 : la segmentation en couches, les champs issus de modèles de mouvement estimés (amplitude des vecteurs multipliée par 2 par rapport aux déplacements calculés), et les images de différence recalées.

Chapitre F

Étude de la limite de débruitage du filtre temporel récursif avec compensation des mouvements

F.1 En situation monocouche

Réglages. Nous considérons le filtre :

$$\hat{I}(\mathbf{p}, t) = \alpha(\mathbf{p}, t)I(\mathbf{p}, t) + (1 - \alpha(\mathbf{p}, t))I(\mathbf{p} + \mathbf{u}, t - 1), \quad (\text{F.1})$$

ce qui se traduit en terme de variance, sous l'hypothèse d'indépendance temporelle et spatiale¹ du bruit, par :

$$\sigma_{\hat{I}}^2(\mathbf{p}, t) = (1 - \alpha(\mathbf{p}, t))^2 \sigma_I^2(\mathbf{p}, t - 1) + \alpha(\mathbf{p}, t)^2 \sigma(\mathbf{p}, t)^2 \quad (\text{F.2})$$

avec $\sigma(\mathbf{p}, t)$ l'écart-type du bruit de mesure et $\sigma_{\hat{I}}(\mathbf{p}, t)$ l'écart-type du bruit résiduel sur l'image débruitée. Nous choisissons le gain $\alpha(\mathbf{p}, t)$ de sorte à minimiser le bruit de l'image restaurée $\sigma_{\hat{I}}^2(\mathbf{p}, t)$:

$$\frac{d\sigma_{\hat{I}}^2(\mathbf{p}, t)}{d\alpha(\mathbf{p}, t)} = 2 \left(\alpha(\mathbf{p}, t) \sigma^2(\mathbf{p}, t) + (\alpha(\mathbf{p}, t) - 1) \sigma_{\hat{I}}^2(\mathbf{p}, t - 1) \right) = 0 \quad (\text{F.3})$$

Ce qui conduit à choisir :

$$\alpha(\mathbf{p}, t) = \frac{\sigma_{\hat{I}}^2(\mathbf{p}, t - 1)}{\sigma^2(\mathbf{p}, t) + \sigma_{\hat{I}}^2(\mathbf{p}, t - 1)} \quad (\text{F.4})$$

Nous considérons que le bruit de mesure ne dépend ni du temps ni de l'espace : $\sigma(\mathbf{p}, t) = \sigma$. De même, le bruit des images après traitement ne dépend pas de la position du pixel considéré² : $\sigma_{\hat{I}}(\mathbf{p} + \mathbf{u}, t) = \sigma_{\hat{I}}(t)$. Alors :

$$\alpha(t) = \frac{\sigma_{\hat{I}}^2(t - 1)}{\sigma^2 + \sigma_{\hat{I}}^2(t - 1)} \quad (\text{F.5})$$

¹Ce qui implique de négliger la MTF et l'interpolation.

²Ici encore, les propriétés passe-bas de l'interpolation sont négligées.

Valeurs explicites du pas et de la limite de débruitage. Montrons par récurrence :

$$\alpha(t) = \frac{1}{t+1} \quad (\text{F.6})$$

$$\sigma(t)^2 = \frac{1}{t+1}\sigma^2 \quad (\text{F.7})$$

\mathbf{H}_1 : *Initialisation.*

$$\alpha(1) = \frac{\sigma_I^2(0)}{\sigma^2 + \sigma_I^2(0)} \quad (\text{F.8})$$

$$= \frac{\sigma^2}{2\sigma^2} = \frac{1}{2} \quad (\text{F.9})$$

puisque en fait la première image n'a pas été débruitée. D'autre part,

$$\sigma_I^2(1) = (1 - \alpha(1))^2 \sigma_I^2(0) + \alpha(1)^2 \sigma^2 \quad (\text{F.10})$$

$$= \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 = \frac{1}{2}\sigma^2 \quad (\text{F.11})$$

La propriété est donc démontrée à la date $t = 1$.

$\mathbf{H}_t \rightarrow \mathbf{H}_{t+1}$: *Hérédité.* Supposons la propriété vraie à la date t . Alors, d'après (F.5) :

$$\alpha(t+1) = \frac{\sigma_I^2(t)}{\sigma^2 + \sigma_I^2(t)} \quad (\text{F.12})$$

$$= \frac{\frac{1}{t+1}\sigma^2}{\sigma^2 + \frac{1}{t+1}\sigma^2} \quad (\text{F.13})$$

$$= \frac{\frac{1}{t+1}\sigma^2}{\frac{(t+2)}{t+1}\sigma^2} \quad (\text{F.14})$$

$$= \frac{1}{t+2} \quad (\text{F.15})$$

La relation (F.2) s'écrit alors :

$$\sigma_I^2(t+1) = (1 - \alpha(t+1))^2 \sigma_I^2(t) + \alpha(t+1)^2 \sigma(t)^2 \quad (\text{F.16})$$

$$= \left(1 - \frac{1}{t+2}\right)^2 \frac{\sigma^2}{t+1} + \left(\frac{1}{t+2}\right)^2 \sigma^2 \quad (\text{F.17})$$

$$= \frac{(t+1)\sigma^2}{(t+2)^2} + \frac{\sigma^2}{(t+2)^2} \quad (\text{F.18})$$

$$= \frac{\sigma^2}{(t+2)} \quad (\text{F.19})$$

Ce qui montre la propriété et achève la démonstration. \square

F.2 En situation de bi-transparence

Réglages. Le filtre considéré est le suivant :

$$\begin{aligned} \hat{I}(\mathbf{p}, t) &= \alpha(\mathbf{p}, t)I(\mathbf{p}, t) \\ &+ (1 - \alpha(\mathbf{p}, t))(I(\mathbf{p} + \mathbf{u}, t - 1) + I(\mathbf{p} + \mathbf{v}, t - 1) \\ &\quad - I(\mathbf{p} + \mathbf{u} + \mathbf{v}, t - 2)) \end{aligned} \quad (\text{F.20})$$

Ce qui se traduit pour les variances, sous hypothèse d'indépendance temporelle et spatiale³ du bruit :

$$\begin{aligned} \sigma_I^2(\mathbf{p}, t) &= \alpha(\mathbf{p}, t)^2 \sigma^2(\mathbf{p}, t) \\ &+ (1 - \alpha(\mathbf{p}, t))^2 (\sigma_I^2(\mathbf{p} + \mathbf{u}, t - 1) \\ &\quad + \sigma_I^2(\mathbf{p} + \mathbf{v}, t - 1) + \sigma_I^2(\mathbf{p} + \mathbf{u} + \mathbf{v}, t - 2)) \end{aligned} \quad (\text{F.21})$$

Comme dans le cas monocouche, nous abandonnons la dépendance spatiale des écarts-type, puisque nous ne pouvons de toute façon pas tirer profit d'une modélisation exacte du problème en pratique. Nous ne savons en effet pas estimer localement la valeur des écarts-type du bruit de manière fiable. Dans ces conditions, et comme le filtre que nous étudions n'est pas adaptatif, le gain α ne dépend plus de la position spatiale non plus. Comme enfin le bruit de mesure ne dépend pas du temps, la relation (F.21) devient :

$$\sigma_I^2(t) = \alpha(t)^2 \sigma^2 + (1 - \alpha(t))^2 (2\sigma_I^2(t - 1) + \sigma_I^2(t - 2)) \quad (\text{F.22})$$

Afin de minimiser $\sigma_I^2(t)$, on cherche à annuler la dérivée de (F.22) par rapport à $\alpha(\mathbf{p}, t)$:

$$\frac{d\sigma_I^2(t)}{d\alpha(t)} = 2 \left(\alpha(t) \sigma^2 + (\alpha(t) - 1) (2\sigma_I^2(t - 1) + \sigma_I^2(t - 2)) \right) = 0 \quad (\text{F.23})$$

D'où

$$\alpha(t) = \frac{2\sigma_I^2(t - 1) + \sigma_I^2(t - 2)}{2\sigma_I^2(t - 1) + \sigma_I^2(t - 2) + \sigma^2} \quad (\text{F.24})$$

Points fixes. Le niveau de bruit de l'image après traitement vérifie alors :

$$\sigma_I^2(t) = \frac{(2\sigma_I^2(t - 1) + \sigma_I^2(t - 2))^2 \sigma^2 + \sigma^4 (2\sigma_I^2(t - 1) + \sigma_I^2(t - 2))}{(2\sigma_I^2(t - 1) + \sigma_I^2(t - 2) + \sigma^2)^2} \quad (\text{F.25})$$

Les limites possibles de cette suite sont données par les points fixes : $\bar{\sigma}^2 = \bar{V}$:

$$\bar{V} = f(\bar{V}) = \frac{9\bar{V}^2 \sigma^2 + 3\sigma^4 \bar{V}}{(3\bar{V} + \sigma^2)^2} \quad (\text{F.26})$$

$$\bar{V} (9\bar{V}^2 - 3\sigma^2 \bar{V} - 2\sigma^4) = 0 \quad (\text{F.27})$$

$$9\bar{V} \left(\bar{V} + \frac{1}{3}\sigma^2 \right) \left(\bar{V} - \frac{2}{3}\sigma^2 \right) = 0 \quad (\text{F.28})$$

³Ce qui implique de négliger la MTF et l'interpolation.

Les trois points fixes sont donc $-\frac{1}{3}\sigma^2$, 0 et $\frac{2}{3}\sigma^2$.

Le premier d'entre eux correspond à une variance négative et donc n'a pas de sens physique ici. Pour savoir si les deux autres sont attractifs ou répulsifs, il nous faut étudier la fonction suivante :

$$g(V) = \frac{9\sigma^2\bar{V}^2 + 3\sigma^4\bar{V}}{(3\bar{V} + \sigma^2)^2} - \bar{V} \quad (\text{F.29})$$

$$= \frac{-9\bar{V}^3 + 3\sigma^2\bar{V}^2 + 2\sigma^4\bar{V}}{(3\bar{V} + \sigma^2)^2} \quad (\text{F.30})$$

Plus précisément, si sa dérivée en un point fixe est de norme supérieure à 1, le point est répulsif. Dans le cas contraire, il est attracteur.

$$g'(V) = \frac{(-27V^2 + 6\sigma^2V + 2\sigma^4)(3V + \sigma^2)^2 - 6(-9V^3 + 3\sigma^2V^2 + 2\sigma^4V)(3V + \sigma^2)}{(3V + \sigma^2)^4} \quad (\text{F.31})$$

$$= \frac{27V^3 - 27\sigma^2V^2 + 2\sigma^6}{(3V + \sigma^2)^3} \quad (\text{F.32})$$

Pour les deux points fixes qui nous intéressent, cette fonction vaut :

$$g'(0) = 2 \quad (\text{F.33})$$

$$g'\left(\frac{2}{3}\sigma^2\right) = 0 \quad (\text{F.34})$$

Finalement, même si la suite a deux points fixes, $\frac{2}{3}\sigma^2$ est la seule limite finie possible.

Convergence. La suite pourrait néanmoins diverger. Nous devons étudier comment évolue sa distance au point fixe attracteur. La suite qui nous intéresse est :

$$\hat{V}_2(t) = \sigma_I^2(t) - \frac{2}{3}\sigma^2 \quad (\text{F.35})$$

On passe d'un élément au suivant de cette suite comme suit :

$$\hat{V}_2(t+1) = \sigma_I^2(t+1) - \frac{2}{3}\sigma^2 \quad (\text{F.36})$$

$$= f(\sigma_I^2(t)) - \frac{2}{3}\sigma^2 \quad (\text{F.37})$$

$$= f\left(\hat{V}_2(t) + \frac{2}{3}\sigma^2\right) - \frac{2}{3}\sigma^2 \quad (\text{F.38})$$

$$= \frac{9\sigma^2\hat{V}_2(t)^2 + 15\sigma^4\hat{V}_2(t) + 6\sigma^6}{(3\hat{V}_2(t) + 3\sigma^2)^2} - \frac{2}{3}\sigma^2 \quad (\text{F.39})$$

$$= \frac{3\sigma^2\hat{V}_2(t)^2 + 3\sigma^4\hat{V}_2(t)}{(3\hat{V}_2(t) + 3\sigma^2)^2} \quad (\text{F.40})$$

$$= \frac{\sigma^2}{3} \frac{\hat{V}_2(t)}{\hat{V}_2(t) + \sigma^2} \quad (\text{F.41})$$

$$= \frac{\sigma^2}{3(\hat{V}_2(t) + \sigma^2)} \hat{V}_2(t) \quad (\text{F.42})$$

Le premier terme de cette dernière fonction est strictement inférieur à 1 lorsque $\hat{V}_2(t)$ est strictement supérieur à $-\frac{2}{3}\sigma^2$ (c'est-à-dire pour toutes les valeurs de la variance excepté le point fixe répulsif 0).

La suite des variances converge donc de manière strictement monotone vers le point fixe attracteur $\frac{2}{3}\sigma^2$. Seule la valeur 0 fait exception, mais elle n'a pas de sens dans notre problème : cela revient à dire que si les premières images de la séquence sont non bruitées, nous sommes en mesure de l'intégralité de la séquence sans intégrer une seule des mesures.

F.3 En situation de transparence à n couches

Règlages. Nous nous plaçons maintenant dans le cas général de la transparence à n couches. L'équation (2.39) a pour équivalent (2.42), que nous rappelons ici :

$$I(\mathbf{p}, t) = \sum_{i=t-n}^{t-1} (-1)^{t-i} \sum_{\{j_1, \dots, j_{t-i}\} \in P_{t-i}([1, n])} I(\mathbf{p} + \mathbf{w}_{j_1} \dots + \mathbf{w}_{j_{t-i}}, i) \quad (\text{F.43})$$

où $P_j([1, n])$ note les permutations possibles de j éléments de $[1, n]$. Cette expression n'est exacte que si les mouvements transparents sont constants sur $n - 1$ intervalles de temps successifs.

L'équivalent de (F.20) au cas de n couches s'écrit donc :

$$\begin{aligned} \hat{I}(\mathbf{p}, t) &= \alpha(\mathbf{p}, t)I(\mathbf{p}, t) \\ &+ (1 - \alpha(\mathbf{p}, t)) \\ &\quad \sum_{i=t-n}^{t-1} (-1)^{t-i} \sum_{\{j_1, \dots, j_{t-i}\} \in P_{t-i}([1, n])} I(\mathbf{p} + \mathbf{w}_{j_1} + \dots + \mathbf{w}_{j_{t-i}}, i) \end{aligned} \quad (\text{F.44})$$

Sous des hypothèses d'indépendance temporelle et spatiale du bruit, cela se traduit en terme de variances par :

$$\begin{aligned} \sigma_I^2(t) &= \alpha(t)^2 \sigma^2 \\ &+ (1 - \alpha(t))^2 \sum_{i=t-n}^{t-1} \sum_{\{j_1, j_2, \dots, j_n\} \in P_{t-i}([1, n])} \sigma_I^2(i) \end{aligned} \quad (\text{F.45})$$

$$= \alpha(t)^2 \sigma^2 + (1 - \alpha(t))^2 \sum_{i=t-n}^{t-1} C_n^{t-i} \sigma_I^2(i) \quad (\text{F.46})$$

Pour établir cette dernière équation, nous avons dénombré le nombre de permutations à $t - i$ éléments parmi n . Il y en a $C_n^{t-i} = \frac{n!}{(t-i)!(n-t-i)!}$.

Nous choisissons à chaque instant le pas minimisant (F.46), c'est-à-dire vérifiant :

$$\frac{d\sigma_I^2(t)}{d\alpha(t)} = 2 \left(\alpha(t)\sigma^2 - (1 - \alpha(t)) \sum_{i=t-n}^{t-1} C_n^{t-i} \sigma_I^2(i) \right) \quad (\text{F.47})$$

$$= 0 \quad (\text{F.48})$$

D'où :

$$\alpha(t) = \frac{\sum_{i=t-n}^{t-1} C_n^{t-i} \sigma_f^2(i)}{\sum_{i=t-n}^{t-1} C_n^{t-i} \sigma_f^2(i) + \sigma^2} \quad (\text{F.49})$$

Avec ce choix de pas, (F.46) devient :

$$\sigma_f^2(t) = \frac{(\sum_{i=t-n}^{t-1} C_n^{t-i} \sigma_f^2(i))^2 \sigma^2 + \sigma^4 (\sum_{i=t-n}^{t-1} C_n^{t-i} \sigma_f^2(i))}{(\sum_{i=t-n}^{t-1} C_n^{t-i} \sigma_f^2(i) + \sigma^2)^2} \quad (\text{F.50})$$

Points fixes. Cette suite définit les variances idéales que ce filtre nous permet d'obtenir après traitement de la séquence. Pour étudier son comportement asymptotique, nous devons étudier ses points fixes éventuels. Nous notons ces points fixes $\bar{\sigma}^2 = \bar{V}$:

$$\bar{V} = \frac{(\sum_{i=t-n}^{t-1} C_n^{t-i})^2 \sigma^2 \bar{V}^2 + (\sum_{i=t-n}^{t-1} C_n^{t-i}) \sigma^4 \bar{V}}{(\sum_{i=t-n}^{t-1} C_n^{t-i} \bar{V} + \sigma^2)^2} \quad (\text{F.51})$$

Il est par ailleurs possible de calculer explicitement $\sum_{i=t-n}^{t-1} C_n^{t-i}$, en se basant sur la formule du binôme :

$$2^n = (1+1)^n = \sum_{k=0}^n C_n^k \quad (\text{F.52})$$

$$= \sum_{i=t-n}^t C_n^{t-i} \quad (\text{F.53})$$

$$= \sum_{i=t-n}^{t-1} C_n^{t-i} + 1 \quad (\text{F.54})$$

D'où :

$$\sum_{i=t-n}^{t-1} C_n^{t-i} = 2^n - 1 \quad (\text{F.55})$$

De sorte que (F.51) devient :

$$\bar{V} = f(\bar{V}) = \frac{(2^n - 1)^2 \sigma^2 \bar{V}^2 + (2^n - 1) \sigma^4 \bar{V}}{((2^n - 1) \bar{V} + \sigma^2)^2} \quad (\text{F.56})$$

$$(\text{F.57})$$

Résolvons cette équation :

$$\bar{V} ((2^n - 1)^2 \sigma^2 \bar{V}^2 + (2(2^n - 1) - (2^n - 1)^2) \sigma^2 \bar{V} + (2 - 2^n) \sigma^4) = 0 \quad (\text{F.58})$$

$$(2^n - 1)^2 \sigma^2 \bar{V} (\bar{V} + \frac{1}{2^n - 1} \sigma^2) (\bar{V} - \frac{2^n - 2}{2^n - 1} \sigma^2) = 0 \quad (\text{F.59})$$

Les trois points fixes sont donc $-\frac{1}{2^n - 1} \sigma^2$, 0 et $\frac{2^n - 2}{2^n - 1} \sigma^2$. Si le premier n'a pas de sens en terme de débruitage, nous devons étudier les deux autres pour déterminer s'ils sont

attracteurs ou répulsifs.

Nous formons pour cela la fonction g :

$$g(\bar{V}) = f(\bar{V}) - \bar{V} \quad (\text{F.60})$$

$$= \frac{(2^n - 1)^2 \sigma^2 \bar{V}^2 + (2^n - 1) \sigma^4 \bar{V} - ((2^n - 1) \bar{V} + \sigma^2)^2 \bar{V}}{((2^n - 1) \bar{V} + \sigma^2)^2} \quad (\text{F.61})$$

$$= \frac{-(2^n - 1)^2 \bar{V}^3 + ((2^n - 1)^2 - 2(2^n - 1)) \sigma^2 \bar{V}^2 + (2^n - 2) \sigma^4 \bar{V}}{((2^n - 1) \bar{V} + \sigma^2)^2} \quad (\text{F.62})$$

Sa dérivée vaut :

$$g'(\bar{V}) = \frac{(2^n - 1)^3 \bar{V}^3 - 3(2^n - 1)^2 \sigma^2 \bar{V}^2 + (2^n - 1)(2^n - 4) \sigma^4 \bar{V} + (2^n - 2) \sigma^6}{((2^n - 1) \bar{V} + \sigma^2)^3} \quad (\text{F.63})$$

Pour les deux points fixes considérés, nous avons :

$$g'(0) = (2^n - 1) - 1 \quad (\text{F.64})$$

$$g'\left(\frac{2^n - 2}{2^n - 1} \sigma^2\right) = \frac{(2^n - 2)^3 \sigma^6 - 3(2^n - 2)^2 \sigma^6 + (2^n - 2)(2^n - 4) \sigma^6 + (2^n - 2) \sigma^6}{(\sigma^2 + (2^n - 2) \sigma^2)^3} \quad (\text{F.65})$$

$$= \frac{(2^n - 1)^2 - 5(2^n - 1) + 6}{(2^n - 1)^2} \quad (\text{F.66})$$

$$= \frac{(2^n - 3)(2^n - 4)}{(2^n - 1)^2} \quad (\text{F.67})$$

Dès que le nombre de couches est supérieur ou égal à 2, le point fixe 0 est répulsif. Il n'y a guère que dans le cas monocouche qu'il est attractif (et il fusionne d'ailleurs avec l'autre point fixe).

De même, les deux termes au numérateur de (F.67) étant inférieurs à $(2^n - 1)$, la dérivée de g au point $\frac{2^n - 2}{2^n - 1} \sigma^2$ est toujours de norme inférieure à 1 : il est attractif.

Convergence. La seule limite finie de la suite (F.50) est donc $\frac{2^n - 2}{2^n - 1} \sigma^2$. Cependant, rien ne nous assure qu'elle ne diverge pas. Nous étudions donc le comportement de la suite des variance relativement à ce point fixe, en introduisant la variable intermédiaire $V_2(t) = V(t) - \frac{2^n - 2}{2^n - 1} \sigma^2$:

$$V_2(t+1) = V(t+1) - \frac{2^n - 2}{2^n - 1} \sigma^2 \quad (\text{F.68})$$

$$= \frac{(2^n - 1)^2 \sigma^2 V^2(t) + (2^n - 1) \sigma^4 V(t)}{((2^n - 1)V(t) + \sigma^2)^2} - \frac{2^n - 2}{2^n - 1} \sigma^2 \quad (\text{F.69})$$

$$= \frac{(2^n - 1)^2 \sigma^2 (V_2^2(t) + \frac{2^n - 2}{2^n - 1} \sigma^2) + (2^n - 1) \sigma^4 (V_2(t) + \frac{2^n - 2}{2^n - 1} \sigma^2)}{((2^n - 1)(V_2(t) + \frac{2^n - 2}{2^n - 1} \sigma^2) + \sigma^2)^2} - \frac{2^n - 2}{2^n - 1} \sigma^2 \quad (\text{F.70})$$

$$= \frac{(2^n - 1)^2 \sigma^2 V_2^2(t) + (2(2^n - 1)^2 - (2^n - 1)) \sigma^4 V_2(t) + ((2^n - 1)^2 - (2^n - 1)) \sigma^2}{((2^n - 1)(V_2(t) + \frac{2^n - 2}{2^n - 1} \sigma^2) + \sigma^2)^2} - \frac{2^n - 2}{2^n - 1} \sigma^2 \quad (\text{F.71})$$

$$= \frac{(2^n - 1) \sigma^2 V_2^2(t) + (2^n - 1) \sigma^4 V_2(t)}{(2^n - 1)^2 (\sigma^2 + V_2(t))} \quad (\text{F.72})$$

$$= \frac{\sigma^2}{(2^n - 1)(\sigma^2 + V_2(t))} V_2(t) \quad (\text{F.73})$$

La fonction $\frac{\sigma^2}{(2^n - 1)(\sigma^2 + V_2(t))}$ est décroissante et vaut 1 en $V_2(t) = -\frac{2^n - 2}{2^n - 1} \sigma^2$, soit lorsque $V(t) = 0$. La suite $V_2(t)$ est donc strictement décroissante, ce qui implique que les variances successives $V(t)$ des images débruitées sont de plus en plus proches de l'unique point fixe attracteur $\frac{2^n - 2}{2^n - 1} \sigma^2$.

Le filtrage envisagé mène donc à une limite asymptotique de débruitage $\frac{2^n - 2}{2^n - 1} \sigma^2$. Les valeurs de cette limite pour le nombre de couches $n = 1 \dots 4$ sont reportées sur le tableau F.1. On peut noter une limite de débruitage qui reste élevée dès le cas bi-couche, et qui le devient encore plus dans des situations de transparence multiple.

La seule exception a lieu pour le cas particulier où les images non bruitées sont connues pour les premiers instants de la séquence. Il est alors possible de reconstruire l'intégralité de la séquence non bruitée. Ce cas n'a pas d'intérêt en pratique.

TAB. F.1: Limite asymptotique de débruitage théorique du filtre récursif temporel avec compensation du mouvement dans une situation à n couches. Nous donnons la limite théorique en variance, ainsi qu'en écart-type.

n	1	2	3	4
$\frac{2^n - 2}{2^n - 1}$	0	$2/3 \simeq 0,667$	$6/7 \simeq 0,857$	$14/15 \simeq 0,933$
$\sqrt{\frac{2^n - 2}{2^n - 1}}$	0	0,816	0,926	0.966

Chapitre G

Calcul des coefficients du filtrage hybride

Nous menons dans cette annexe les calculs permettant de spécifier les gains optimaux pour le filtre hybride, dont nous rappelons la formule :

$$\begin{aligned}\hat{I}(\mathbf{p}, t) = & \alpha(\mathbf{p})I(\mathbf{p}, t) + \beta(\mathbf{p})\hat{I}(\mathbf{p} + \mathbf{u}(\mathbf{p}), t - 1) \\ & + \delta(\mathbf{p})\hat{I}(\mathbf{p} + \mathbf{v}(\mathbf{p}), t - 1) + \gamma(\mathbf{p})\bar{I}(\mathbf{p}, t)\end{aligned}\quad (\text{G.1})$$

Nous menons les calculs dans quatre configurations :

- (**H₀**) : Les deux couches sont hétérogènes.
- (**H₁**) : La première couche (celle qui correspond à \mathbf{u}) est hétérogène.
- (**H₂**) : La deuxième couche (celle qui correspond à \mathbf{v}) est hétérogène.
- (**H₃**) : Les deux couches sont homogènes.

G.1 Dans l'hypothèse (H_0)

Nous ne pouvons dans cette hypothèse utiliser ni $\hat{I}(\mathbf{p} + \mathbf{u}(\mathbf{p}), t - 1)$, ni $\hat{I}(\mathbf{p} + \mathbf{v}(\mathbf{p}), t - 1)$. Il nous faut donc fixer $\beta(\mathbf{p}) = 0$ et $\delta(\mathbf{p}) = 0$.

Nous nous retrouvons dans la situation de la compensation de mouvements transparents du chapitre 7. D'après les calculs effectués en 7.2.2 :

$$\alpha(\mathbf{p}, t) = \frac{\sigma_i(\mathbf{p} + \mathbf{u}, t - 1)^2 + \sigma_i(\mathbf{p} + \mathbf{v}, t - 1)^2 + \sigma_i(\mathbf{p} + \mathbf{u} + \mathbf{v}, t - 2)^2}{\sigma_i(\mathbf{p} + \mathbf{u}, t - 1)^2 + \sigma_i(\mathbf{p} + \mathbf{v}, t - 1)^2 + \sigma_i(\mathbf{p} + \mathbf{u} + \mathbf{v}, t - 2)^2 + \sigma(\mathbf{p}, t)^2} \quad (\text{G.2})$$

$$\gamma(\mathbf{p}, t) = \frac{\sigma(\mathbf{p}, t)^2}{\sigma_i(\mathbf{p} + \mathbf{u}, t - 1)^2 + \sigma_i(\mathbf{p} + \mathbf{v}, t - 1)^2 + \sigma_i(\mathbf{p} + \mathbf{u} + \mathbf{v}, t - 2)^2 + \sigma(\mathbf{p}, t)^2} \quad (\text{G.3})$$

Dans toute cette annexe, nous abandonnons la dépendance spatiale des écarts-type, puisque nous ne pouvons de toute façon pas tirer profit d'une modélisation exacte du problème en pratique. Nous ne savons en effet pas estimer localement la valeur des écarts-type du bruit de manière fiable. Dans ces conditions, et comme le filtre que nous étudions n'est pas adaptatif, les facteurs calculés ne dépendent plus de la position spatiale non plus. Comme

enfin le bruit de mesure ne dépend pas du temps, les facteurs α et γ deviennent :

$$\alpha(t) = \frac{2\sigma_{\hat{I}}(t-1)^2 + \sigma_{\hat{I}}(t-2)^2}{2\sigma_{\hat{I}}(t-1)^2 + \sigma_{\hat{I}}(t-2)^2 + \sigma^2} \quad (\text{G.4})$$

$$\gamma(t) = \frac{\sigma^2}{2\sigma_{\hat{I}}(t-1)^2 + \sigma_{\hat{I}}(t-2)^2 + \sigma^2} \quad (\text{G.5})$$

G.2 Cas de l'hypothèse (H_1)

Dans cette hypothèse, il faut s'interdire d'utiliser $\hat{I}(\mathbf{p} + \mathbf{v}(\mathbf{p}), t-1)$, et donc fixer $\delta(\mathbf{p}) = 0$.

L'équation (G.1) se traduit sur les variances des différentes images comme suit¹ :

$$\sigma_{\hat{I}}^2(t) = \alpha(t)^2\sigma^2 + \beta(t)^2\sigma_{\hat{I}}^2(t-1) + \gamma(t)^2\bar{\sigma}^2(t) \quad (\text{G.6})$$

avec d'autre part :

$$\alpha(t) + \beta(t) + \gamma(t) + \delta(t) = 1 \quad (\text{G.7})$$

Nous reprenons (G.6) en allégeant les notations :

$$\sigma_{\hat{I}}^2(t) = \alpha^2(t)\sigma^2 + \beta^2(t)\sigma_{\hat{I}}^2(t-1) + \gamma^2(t)\bar{\sigma}^2(t) \quad (\text{G.8})$$

$$\sigma_{\hat{I}}^2(t) = \alpha^2(t)\sigma^2 + \beta^2(t)\sigma_{\hat{I}}^2(t-1) + (1 - \alpha(t) - \beta(t))^2\bar{\sigma}^2(t) \quad (\text{G.9})$$

Nous souhaitons minimiser (G.9) par rapport aux gains :

$$\frac{\partial\sigma_{\hat{I}}^2(t)}{\partial\alpha(t)} = 2(\alpha(t)(\sigma^2 + \bar{\sigma}^2(t)) + \beta(t)\bar{\sigma}^2(t) - \bar{\sigma}^2(t)) = 0 \quad (\text{G.10})$$

$$\frac{\partial\sigma_{\hat{I}}^2(t)}{\partial\beta(t)} = 2(\alpha(t)\bar{\sigma}^2(t) + \beta(t)(\sigma_{\hat{I}}^2(t-1) + \bar{\sigma}^2(t)) - \bar{\sigma}^2(t)) = 0 \quad (\text{G.11})$$

(G.10)-(G.11) nous conduit à :

$$\beta(t) = \frac{\sigma^2}{\sigma_{\hat{I}}^2(t-1)}\alpha(t) \quad (\text{G.12})$$

En réinjectant cette équation dans (G.10), et en multipliant le tout par $\sigma_{\hat{I}}^2(t-1)$:

$$\alpha(t)(\sigma^2\sigma_{\hat{I}}^2(t-1) + \bar{\sigma}^2(t)\sigma_{\hat{I}}^2(t-1) + \bar{\sigma}^2(t)\sigma^2) = \bar{\sigma}^2(t)\sigma_{\hat{I}}^2(t-1) \quad (\text{G.13})$$

D'où l'on tire finalement :

$$\alpha(t) = \frac{\bar{\sigma}^2(t)\sigma_{\hat{I}}^2(t-1)}{\sigma^2\sigma_{\hat{I}}^2(t-1) + \bar{\sigma}^2(t)\sigma_{\hat{I}}^2(t-1) + \bar{\sigma}^2(t)\sigma^2} \quad (\text{G.14})$$

$$\beta(t) = \frac{\bar{\sigma}^2(t)\sigma^2}{\sigma^2\sigma_{\hat{I}}^2(t-1) + \bar{\sigma}^2(t)\sigma_{\hat{I}}^2(t-1) + \bar{\sigma}^2(t)\sigma^2} \quad (\text{G.15})$$

$$\delta(t) = 0 \quad (\text{G.16})$$

$$\gamma(t) = \frac{\sigma_{\hat{I}}^2(t-1)\sigma^2}{\sigma^2\sigma_{\hat{I}}^2(t-1) + \bar{\sigma}^2(t)\sigma_{\hat{I}}^2(t-1) + \bar{\sigma}^2(t)\sigma^2} \quad (\text{G.17})$$

¹Sous l'hypothèse d'indépendance des bruits. Celle-ci n'est toutefois pas exacte à cause de l'interpolation et de la *MTF*, qui impliquent une corrélation spatiale.

G.3 Cas de l'hypothèse (H_2)

Ce cas est exactement symétrique au précédent :

$$\alpha(t) = \frac{\bar{\sigma}^2(t)\sigma_{\bar{I}}^2(t-1)}{\sigma^2\sigma_{\bar{I}}^2(t-1) + \bar{\sigma}^2(t)\sigma_{\bar{I}}^2(t-1) + \bar{\sigma}^2(t)\sigma^2} \quad (\text{G.18})$$

$$\beta(t) = 0 \quad (\text{G.19})$$

$$\delta(t) = \frac{\bar{\sigma}^2(t)\sigma^2}{\sigma^2\sigma_{\bar{I}}^2(t-1) + \bar{\sigma}^2(t)\sigma_{\bar{I}}^2(t-1) + \bar{\sigma}^2(t)\sigma^2} \quad (\text{G.20})$$

$$\gamma(t) = \frac{\sigma_{\bar{I}}^2(t-1)\sigma^2}{\sigma^2\sigma_{\bar{I}}^2(t-1) + \bar{\sigma}^2(t)\sigma_{\bar{I}}^2(t-1) + \bar{\sigma}^2(t)\sigma^2} \quad (\text{G.21})$$

G.4 Cas de l'hypothèse (H_3)

Dans ce dernier cas, les quatre images entrent en ligne de compte. L'équation (G.1) se traduit sur les variances des différentes images comme suit² :

$$\sigma_{\bar{I}}^2(t) = \alpha^2(t)\sigma^2 + \beta^2(t) + \delta^2(t)\sigma_{\bar{I}}^2(t-1) + \gamma^2(t)\bar{\sigma}^2(t) \quad (\text{G.22})$$

$$\sigma_{\bar{I}}^2(t) = \alpha^2(t)\sigma^2 + \beta^2(t) + \delta^2(t)\sigma_{\bar{I}}^2(t-1) + (1 - \alpha(t) - \beta(t) - \delta(t))^2\bar{\sigma}^2(t) \quad (\text{G.23})$$

Nous dérivons cette expression par rapport à ces trois inconnues indépendantes pour la minimiser :

$$\frac{\partial\sigma_{\bar{I}}^2(t)}{\partial\alpha} = 2(\alpha(t)(\sigma^2 + \bar{\sigma}^2(t)) + \beta(t)\bar{\sigma}^2(t) + \delta(t)\bar{\sigma}^2(t) - \bar{\sigma}^2(t)) = 0 \quad (\text{G.24})$$

$$\frac{\partial\sigma_{\bar{I}}^2(t)}{\partial\beta} = 2(\alpha(t)\bar{\sigma}^2(t) + \beta(t)(\sigma_{\bar{I}}^2(t-1) + \bar{\sigma}^2(t)) + \delta\bar{\sigma}^2(t) - \bar{\sigma}^2(t)) = 0 \quad (\text{G.25})$$

$$\frac{\partial\sigma_{\bar{I}}^2(t)}{\partial\delta} = 2(\alpha(t)\bar{\sigma}^2(t) + \beta(t)\bar{\sigma}^2(t) + \delta(t)(\sigma_{\bar{I}}^2(t-1) + \bar{\sigma}^2(t)) - \bar{\sigma}^2(t)) = 0 \quad (\text{G.26})$$

Calculer (G.24) - (G.25) et (G.24) - (G.26) mène à :

$$\beta(t) = \frac{\sigma^2}{\sigma_{\bar{I}}^2(t-1)}\alpha(t) \quad (\text{G.27})$$

$$\gamma(t) = \frac{\sigma^2}{\sigma_{\bar{I}}^2(t-1)}\alpha(t) \quad (\text{G.28})$$

$$(\text{G.29})$$

En réinjectant cette équation dans (G.24), et en multipliant le tout par $\sigma_{\bar{I}}^4(t-1)$:

$$\begin{aligned} \alpha(t)(\sigma^2\sigma_{\bar{I}}^4(t-1) + \bar{\sigma}^2(t)\sigma_{\bar{I}}^4(t-1) + \bar{\sigma}^2(t)\sigma^2\sigma_{\bar{I}}^2(t-1) + \bar{\sigma}^2(t)\sigma^2\sigma_{\bar{I}}^2(t-1)) \\ = \bar{\sigma}^2(t)\sigma_{\bar{I}}^4(t-1) \end{aligned} \quad (\text{G.30})$$

²Sous l'hypothèse d'indépendance des bruits...

Ainsi,

$$\alpha(t) = \frac{\bar{\sigma}^2(t)\sigma_{\hat{I}}^4(t-1)}{\sigma^2\sigma_{\hat{I}}^4(t-1) + \bar{\sigma}^2(t)\sigma_{\hat{I}}^4(t-1) + \bar{\sigma}^2(t)\sigma^2\sigma_{\hat{I}}^2(t-1) + \bar{\sigma}^2(t)\sigma^2\sigma_{\hat{I}}^2(t-1)} \quad (\text{G.31})$$

$$\beta(t) = \frac{\bar{\sigma}^2(t)\sigma_{\hat{I}}^2(t-1)\sigma^2}{\sigma^2\sigma_{\hat{I}}^4(t-1) + \bar{\sigma}^2(t)\sigma_{\hat{I}}^4(t-1) + \bar{\sigma}^2(t)\sigma^2\sigma_{\hat{I}}^2(t-1) + \bar{\sigma}^2(t)\sigma^2\sigma_{\hat{I}}^2(t-1)} \quad (\text{G.32})$$

$$\delta(t) = \frac{\bar{\sigma}^2(t)\sigma_{\hat{I}}^2(t-1)\sigma^2}{\sigma^2\sigma_{\hat{I}}^4(t-1) + \bar{\sigma}^2(t)\sigma_{\hat{I}}^4(t-1) + \bar{\sigma}^2(t)\sigma^2\sigma_{\hat{I}}^2(t-1) + \bar{\sigma}^2(t)\sigma^2\sigma_{\hat{I}}^2(t-1)} \quad (\text{G.33})$$

$$\gamma(t) = \frac{\sigma_{\hat{I}}^4(t-1)\sigma^2}{\sigma^2\sigma_{\hat{I}}^4(t-1) + \bar{\sigma}^2(t)\sigma_{\hat{I}}^4(t-1) + \bar{\sigma}^2(t)\sigma^2\sigma_{\hat{I}}^2(t-1) + \bar{\sigma}^2(t)\sigma^2\sigma_{\hat{I}}^2(t-1)} \quad (\text{G.34})$$

L'ensemble de ces résultats sont rassemblés dans le tableau 8.1.

Liste des algorithmes

1	Algorithme d'estimation de mouvements transparents en situation de bi-transparence.	83
2	Algorithme d'initialisation de l'estimation de mouvement transparent en situation de bi-transparence distribuée à segmentation donnée	112
3	Algorithme du schéma joint d'estimation de mouvements paramétriques transparents et de segmentation en couches.	129
4	Algorithme de l'initialisation de l'estimation de mouvement transparent en situation de bi-transparence distribuée sans <i>a-priori</i>	131
5	Algorithme du filtre hybride adaptatif.	189
6	Algorithme de minimisation de l'équation de contrainte du mouvement avec transparence linéarisée au premier ordre.	233
7	Algorithme de minimisation associé à l'équation contrainte du mouvement avec la transparence linéarisée au second ordre.	238

Bibliographie

- [Aach 96a] T. Aach & D. Kunz. *Anisotropic spectral magnitude estimation filters for noise reduction and image enhancement*. In Proc. International Conference on Image Processing, pages 335–338, Lausanne, septembre 1996.
- [Aach 96b] T. Aach, D. Kunz, R. Florent & S. Makram-Ebeid. *Noise reduction and image enhancement algorithms for low-dose X-Ray fluoroscopy*. In Bildverarbeitung für die Medizin, 1996.
- [Aach 99] T. Aach, U. Schiebel & G. Spekowius. *Digital image acquisition and processing in x-ray imaging*. Journal of electronic imaging, vol. 8, pages 7–22, 1999.
- [Aach 00] T. Aach & D. Kunz. *A lapped directional transform for spectral image analysis and its application to restoration and enhancement*. Signal Processing, vol. 80, no. 11, pages 2347–2364, 2000.
- [Alp 91] M.B. Alp & Y. Neuvo. *3-dimensional median filters for image sequence processing*. In IEEE Proc. Int. Conf. Acoust., Speech and Signal Processing, pages 2917–2920, Toronto, Canada, mai 1991.
- [Anscombe 48] F.J. Anscombe. *The transformation of Poisson, binomial and negative-binomial data*. Biometrika, vol. 35, pages 246–254, 1948.
- [Arse 91] G.R. Arse. *Multistage order statistic filters for image sequence processing*. IEEE Trans on Signal Processing, vol. SP-39, pages 1147–1163, mai 1991.
- [Auvray 05a] V. Auvray, J. Liénard & P. Bouthemy. *Estimation paramétrique multirésolution de mouvements transparents*. In GRETSI 2005, Louvain-la-Neuve, septembre 2005.
- [Auvray 05b] V. Auvray, J. Liénard & P. Bouthemy. *Multiresolution parametric estimation of transparent motions*. In Proc. Int. Conf. on Image Processing (ICIP'05), Genova, septembre 2005.
- [Auvray 05c] V. Auvray, J. Liénard & P. Bouthemy. *Multiresolution parametric estimation of transparent motions and denoising of fluoroscopic images*. In MICCAI'05, volume 2, pages 352–359, Palm Springs, USA, octobre 2005.
- [Auvray 06a] V. Auvray, P. Bouthemy & J. Liénard. *Motion-based segmentation of transparent layers in video sequences*. In Proc. of International

- Workshop of Multimedia Content Representation, Classification and Security, Istanbul, Turkey, sept 2006.
- [Auvray 06b] V. Auvray, P. Bouthemy & J. Liénard. *Motion estimation in X-Ray image sequences with bi-distributed transparency*. In Proc. Int. Conf. on Image Processing (ICIP'06), Atlanta, octobre 2006.
- [Azzabou 06] N. Azzabou, N. Paragios & F. Guichard. *Random walks, constrained multi[ple hypothesis testing and image enhancement*. In European Conference on Computer Vision, ECCV'06, Zurich, juin 2006.
- [Baert 03] S.A.M. Baert, M.A. Viergever & W.J. Niessen. *Guide wire tracking during endovascular interventions*. IEEE Trans. Med. Imaging, vol. 22, no. 8, pages 965–972, 2003.
- [Barth 03] E. Barth, I. Stuke, T. Aach & C. Mota. *Spatio-temporal motion estimation for transparency and occlusions*. In Proceedings of IEEE International Conference on Image Processing (ICIP 03), volume 3, pages 69–72, Barcelone, septembre 2003.
- [Bergen 92a] J.R. Bergen, P. Anandan, K.J. Hanna & R. Hingorani. *Hierarchical model-based motion estimation*. In ECCV'92, volume 588 of *Lecture Notes in Computer Science*, pages 237–285, Santa Margherita Ligure, Italie, mai 1992.
- [Bergen 92b] J.R. Bergen, P.J. Burt, R. Hingorani & S. Peleg. *A three frame algorithm for estimating two-component image motion*. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 14, pages 886–896, décembre 1992.
- [Bertero 88] M. Bertero, T.A. Poggio & V. Torre. *Ill-posed problems in early vision*. Proceedings of the IEEE, vol. 76, no. 869-889, août 1988.
- [Black 93] M.J. Black & P. Anandan. *A framework for the robust estimation of optical flow*. In Fourth International Conf. on Computer Vision, ICCV-93, pages 231–236, Berlin, 1993.
- [Black 96a] M. Black & P. Anandan. *The robust estimation of multiple motions : parametric and piecewise-smooth flow field*. Computer Vision and Image Understanding, vol. 19, no. 1, pages 57–91, janvier 1996.
- [Black 96b] M. Black & A. Rangarajan. *On the unification of line processes, outlier rejection, and robust statistics with applications in early vision*. International Journal of Computer Vision, vol. 19, no. 1, pages 57–92, juillet 1996.
- [Boo 98] K.G. Boo & N.K. Bose. *A motion-compensated spatio-temporal filter for image sequences for signal-dependent noise*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 8, no. 3, pages 287–298, juin 1998.
- [Boulanger 06a] J. Boulanger. *Analyse de séquences d'images de vidéo-microscopie*. These, Univ Rennes 1, 2006.

- [Boulanger 06b] J. Boulanger, Ch. Kervrann & P. Bouthemy. *Adaptive space-time patch-based method for image sequence restoration*. In Proc. Int. Workshop on Statistical Methods in Multi-Image and Video Processing (SMVP'06), Graz, Autriche, mai 2006.
- [Boulanger 06c] J. Boulanger, K. Kervrann & P. Bouthemy. *Space-time adaptation for patch based image sequence restoration*. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2006.
- [Bourdon 04] P. Bourdon, B. Augereau, C. Oliver & C. Chatellier. *Débruitage de séquences d'images couleur par diffusion anisotropes spatiale et temporelle couplées*. In Reconnaissance des Formes et Intelligence Artificielle, RFIA'04, Toulouse, 2004.
- [Bouthemy 93] P. Bouthemy & E. François. *Motion segmentation and qualitative dynamic scene analysis from an image sequence*. International Journal of Computer Vision, vol. 10, no. 2, pages 157–182, avril 1993.
- [Brailean 95a] J.C. Brailean & A.K. Katsaggelos. *Simultaneous recursive displacement estimation and restoration of noisy-blurred image sequences*. IEEE Transactions on Image Processing, vol. 4, no. 9, pages 1236–1251, septembre 1995.
- [Brailean 95b] J.C. Brailean, R. Kleihorst, S. Efstratiadis, A. Katsaggelos & R. Lagendijk. *Noise reduction filters for dynamic image sequences : a review*. Proceedings of the IEEE, vol. 83, no. 9, pages 1272–1292, sept 1995.
- [Buades 05] A. Buades, B. Coll & J.M. Morel. *A review of image denoising algorithms, with a new one*. Multiscale Modeling and Simulation (SIAM interdisciplinary journal), vol. 4, no. 2, pages 490–430, 2005.
- [Cano 83] D. Cano & M. Bénard. *Image sequence processing and dynamic scene analysis*. In 3D Kalman filtering of image sequences, volume F2 of NATO ASI Series. Springer Verlag, 1983.
- [Crawford 82] D.I. Crawford. *Spatio-temporal prefiltering for a video conference coder*. In Proc. Int. IEE Conf. on Electronic Image Processing, pages 236–242, York, UK, juillet 1982.
- [Darrell 91] T. Darrell & A. Pentland. *Robust estimation of a multi-Layer motion representation*. In Proc. IEEE Workshop on Visual Motion, Princeton, New-Jersey, 1991.
- [Darrell 93] T. Darrell & E. Simoncelli. *Separation of transparent motion into layers using velocity-tuned mechanisms*. Rapport technique 244, MIT Media Laboratory Vision and Modelling group, octobre 1993.
- [Dekeyser 00] F. Dekeyser, P. Bouthemy & P. Pérez. *Spatio-temporal Wiener filtering of image sequences using a parametric motion model*. In IEEE Int. Conf. on Image Processing, Vancouver, 2000.
- [Dekker 85] C.B. Dekker, A.J.E.M. Jansen & P.j. van Oterloo. *The contour plot method for noise reduction in digital video*. Acta Electronica, vol. 27, no. 1-2, pages 119–131, 1985.

- [Donoho 94] D.L. Donoho & I.M. Johnstone. *Ideal spatial adaptation via wavelet shrinkage*. Biometrika, vol. 61, pages 425–455, décembre 1994.
- [Dubois 84] E. Dubois & S. Sabri. *Noise reduction in image sequences using motion compensated temporal filtering*. IEEE Trans. Communications, vol. COM-32, pages 826–831, juillet 1984.
- [Dugad 01] R. Dugad & N. Ahuja. *Noise reduction in video by joint spatial and temporal processing*. IEEE Trans. on Circuits and Systems for Video Technology, vol. 4, pages 461–474, septembre 2001.
- [Erdem 92] A.T. Erdem, M.I. Sezan & M.K. Özkan. *Motion compensated multiframe Wiener restoration of blurred and noisy image sequences*. In IEEE Proceedings Int. Conf. Acoust., Speech and Signal Proc., volume III, pages 293–296, San-Francisco, California, USA, 1992.
- [Fan 99] C. Fan & N.M. Namazi. *Simultaneous motion estimation and filtering of image sequences*. IEEE Transactions on Image Processing, vol. 8, no. 12, pages 1788–1795, décembre 1999.
- [Geman 92] S. Geman & D.E. McClure. *A nonlinear filter for film restoration and other problems in image processing*. CVGIP : Graphical models and image processing, vol. 54, pages 281–289, juillet 1992.
- [Hemmendorff 01] M. Hemmendorff. *Motion estimation and compensation in medical imaging, N° 703*. These, Linköping Universitet, Suède, 2001.
- [Holland 77] P.W. Holland & R.E. Welsch. *Robust regression using iteratively reweighted least-squares*. Communication Statistic-Theory Method A, vol. 6, no. 9, pages 813–827, 1977.
- [Hong 89] L. Hong & D. Brzakovic. *Bayesian restoration of image sequences using 3D Markov random fields*. In IEEE Proc. Int. Conf. Acoust., Speech and Signal Processing, volume M1.5, pages 1413–1416, Glasgow, Écosse, mai 1989.
- [Huang 81] T.S. Huang & Y.P.Hsu. Image sequence analysis, chapitre 4, Image sequence enhancement, pages 289–309. Springer Verlag, 1981.
- [Hubert 81] P.J. Hubert. Robust statistics. John Wiley and Sons, 1981.
- [Illingworth 88] J. Illingworth & J. Kittler. *A survey of the Hough transform*. Computer vision, Graphics and Image Processing, vol. 44, pages 87–116, 1988.
- [Irani 93] M. Irani & S. Peleg. *Motion analysis for image enhancement : resolution, occlusion, and transparency*. Journal of Visual Communication and Image Representation, vol. 4, no. 4, pages 324–335, décembre 1993.
- [Irani 94] M. Irani, B. Rousso & S. Peleg. *Computing occluding and transparent motions*. International Journal of Computer Vision, vol. 12, no. 1, pages 5–16, 1994.
- [Kalivas 90] D.S. Kalivas & A.A. Sawchuk. *Motion-compensated enhancement of noisy image sequences*. In Proceedings of the IEEE Conference

- on Acoustics, Speech and Signal Processing, pages 2121–2124, Albuquerque, 1990.
- [Katsaggelos 89] A.K. Katsaggelos, J.N. Driessen, S.N. Efstratiadis & R.L. Lagendijk. *Temporal motion compensated noise filtering of image sequences*. In SPIE Proc. Vis. Comm. and Image Processing, pages 61–70, Boston, USA, novembre 1989.
- [Kervrann 06] C. Kervrann & J. Boulanger. *Optimal spatial adaptation for patch-based image denoising*. IEEE Trans. on image Processing, 2006. octobre.
- [Kleihorst 94] R.P. Kleihorst. *Noise filtering of image sequences*. These, Technische Universiteit Delt, 1994.
- [Kokaram 98] A.C. Kokaram. Motion picture restoration, chapitre 10. Springer, 1998.
- [Kornprobst 98] P. Kornprobst. *Contribution à la restauration d'images et à l'analyse de séquences : Approches Variationnelles et Solutions de Viscosité*. These, Université de Nice-Sophia Antipolis - Faculté des Sciences, Nice-Sophia Antipolis, novembre 1998.
- [Lee 98] Suk Ho Lee & Moon Gi Kang. *Spatio-temporal Video Filtering Algorithm based on 3-D Anisotropic Diffusion Equation*. In International Conference on Image Processing, ICIP'98, volume 3(2), pages 447–450, Chicago, octobre 1998.
- [Lee 06] Sang-Chul Lee & P. Bajcsy. *Intensity correction of fluorescent confocal laser scanning microscope images by mean-weight filtering*. Journal of Microscopy, vol. 221, no. 2, page 122, February 2006.
- [Levin 02] A. Levin, A. Zomet & Y. Weiss. *Learning to perceive transparency from the statistics of natural scenes*. In Neural Information Processing Systems NIPS'02, Vancouver, Canada, décembre 2002.
- [Levin 04a] A. Levin, G. Sommer & Y. Weiss. *Separating reflections from a single image using local features*. In IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'04), pages 306–313, Washington DC, juin 2004.
- [Levin 04b] A. Levin & Y. Weiss. *User-assisted separation of reflections from a single image using a sparsity prior*. In Proc. of the European Conference on Computer Vision (ECCV'04), volume 1, pages 602–613, Prague, mai 2004.
- [Macovski 04] A. Macovski. Medical imaging systems. Prentice-Hall, 2nd edition, 2004.
- [Mallat 98] S. Mallat. A wavelet tour of signal processing. Academic Press, 1998.
- [McMann 78] R.H. McMann. *Digital noise recover for encoded NTSC signals*. SMPTE journal, vol. 97, pages 129–133, mars 1978.
- [Meijering 99] E.H.W. Meijering, K.J. Zuiderveld & M.A. Viergever. *Image registration for digital subtraction angiography*. International Journal of Computer Vision, vol. 31, pages 227–246, 1999.

- [Mikula 04] M. Mikula, T. Preusser & M. Rumpf. *Morphological image sequence processing*. Computing and Visualization in Science, vol. 6, no. 4, pages 197–209, 2004.
- [Milanfar 96] P. Milanfar. *Projection-based frequency-domain estimation of superimposed translational motions*. Journal of the Optical Society of America : A, Optics and Image Science, vol. 13, no. 11, pages 2151–2162, novembre 1996.
- [Milanfar 99] P. Milanfar. *Two-dimensional matched filtering for motion estimation*. IEEE Transactions on Image Processing, vol. 8, no. 3, pages 438–444, mars 1999.
- [Mittal 04] A. Mittal & N. Paragios. *Motion-based background subtraction using adaptive kernel density estimation*. In Computer Vision and Pattern Recognition, 2004. CVPR'04, volume 2, pages 302–309, Washinton DC, juin 2004.
- [Mota 01] C. Mota, I. Stuke & E. Barth. *Analytic solutions for multiple motions*. In Proceedings of the IEEE International Conference on Image Processing (ICIP 01), volume 2, pages 917–920, Thessalonique, octobre 2001.
- [Mota 05a] C. Mota, I. Stuke, T. Aach & E. Barth. *Divide-and-conquer strategies for estimating multiple transparent motions*. In Springer, editeur, Proceedings of the 1st International Workshop on Complex Motion, Lecture Notes on Computer Science, LNCS Vol. 3417, Schloss Reichartshausen, janvier 2005.
- [Mota 05b] C. Mota, I. Stuke, T. Aach & E. Barth. *Spatial motion estimation for transparency and occlusions*. Signal Processing : Image Communication. Elsevier Science, vol. 20-6, pages 529–536, juin 2005.
- [Murray 87] D.W. Murray & H. Buxton. *Scene segmentation from visual motion using global optimization*. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 9, no. 2, pages 220–228, mars 1987.
- [Naqvi 86] S.S.H. Naqvi, N.C. Gallagher & E.J. Coyle. *An application of median filter in digital television*. In IEEE Proc. Int. Acoust., Speech, Signal Processing, pages 2451–2454, Tokyo, Japan, 1986.
- [Nzomigni 93] V. Nzomigni, C. Labit & J. Liénard. *Motion-compensated lossless compression schemes for biomedical sequence storage*. In Picture Coding Symposium, Lausanne, Suisse, March 1993.
- [Odobez 95] J.-M. Odobez & P. Bouthemy. *Robust multiresolution estimation of parametric motion models*. Journal of Visual Communication and Image Representation, vol. 6, no. 4, pages 348–365, décembre 1995.
- [Odobez 98] J.-M. Odobez & P. Bouthemy. *Direct incremental model-based image motion segmentation for video analysis*. Signal Processing, vol. 6, no. 2, pages 143–155, 1998.
- [Özkan 91] M.K. Özkan, M.I. Sezan & A.T. Erdem. *LMMSE restoration of blurred and noisy images*. In Proceedings SPIE Conference on Visual

- Communication and Image Processing, volume 1606, pages 743–754, Boston, 1991.
- [Özkan 93] M.K. Özkan, M.I. Sezan & A.M. Tekalp. *Adaptive motion compensated filtering of noisy image sequences*. IEEE Transactions on Circuits and Systems for Video Technology, vol. CSVT-3, pages 277–290, août 1993.
- [Paragios 98] N. Paragios. *A level set approach for shape-driven segmentation and tracking of the left ventricle*. IEEE Trans. on Medical Imaging, vol. 22, no. 6, pages 773–776, juin 1998.
- [Paragios 05] N. Paragios & R. Deriche. *Geodesic active regions and level set methods for motion estimation and tracking*. Computer Vision and Image Understanding, vol. 97, pages 259–282, 2005.
- [Perona 90] P. Perona & J. Malik. *Scale-space and edge detection using anisotropic diffusion*. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 12, no. 7, pages 629–639, juillet 1990.
- [Pingault 02a] M. Pingault & D. Pellerin. *Motion transparency constraint equation based on a wavelet function decomposition*. In IEEE International Conference on Multimedia and Expo, ICME'2002, Lausanne, août 2002.
- [Pingault 02b] M. Pingault & D. Pellerin. *Optical flow constraint equation extended to transparency*. In XI European Signal Processing Conference EUSIPCO'2002, Toulouse, septembre 2002.
- [Pingault 02c] M. Pingault, D. Pellerin & J. Shimahara. *Estimation du mouvement d'objets transparents dans le domaine fréquentiel*. In Congrès Reconnaissance de Formes et Intelligence Artificielle, RFIA'2002, Angers, janvier 2002.
- [Pingault 03a] M. Pingault. *Estimations fréquentielle et temporelle du mouvement en transparence additive dans les séquences d'images*. These, Laboratoire des Images et des Signaux (LIS), Université Joseph Fourier, Grenoble, octobre 2003.
- [Pingault 03b] M. Pingault, E. Bruno & D. Pellerin. *A robust multiscale B-spline function decomposition for estimating motion transparency*. IEEE Trans. on Image Processing, vol. 12, no. 11, pages 1416–1426, novembre 2003.
- [Pingault 04] M. Pingault & D. Pellerin. *Motion estimation of transparent objects in the frequency domain*. Signal Processing, vol. 84, no. 4, pages 709–719, avril 2004.
- [Pizurica 03] A. Pizurica, V. Zlokolica & W. Philips. *Noise reduction in video sequences using wavelet-domain and temporal filtering*. In SPIE Conference Wavelet Applications in Industrial Processing, Providence, Rhode Island, USA, octobre 2003.
- [Polak 71] E. Polak. *Computational methods in optimization*. New-York :Academic Press, 1971.

- [Press 92] W. Press & al. Numerical recipes in C : The art of scientific computing, chapitre Downhill simplex method in multidimensions, pages 408–412. Cambridge University Press, 1988-1992.
- [Rabbani 87] M. Rabbani, R. Shaw & R. van Mettler. *Detective quantum efficiency of imaging systems with amplifying and scattering mechanisms*. Journal of optical society of America, vol. 4, no. 5, pages 895–901, 1987.
- [Sarel 04] B. Sarel & M. Irani. *Separating transparent layers through layer information exchange*. In European Conference on Computer Vision (ECCV'04), volume 4, pages 328–341, Prague, mai 2004.
- [Sarel 05] B. Sarel & M. Irani. *Separating transparent layers of repetitive dynamic behaviors*. In IEEE International Conference on Computer Vision (ICCV'05), volume 1, pages 26–32, Beijing, 2005.
- [Selesnick 03] I.W. Selesnick & K.Y. Li. *Video denoising using 2D and 3D dual-tree complex wavelet transforms*. Wavelet Applications in Signal and Image Processing X (Proc. SPIE 5207), août 2003.
- [Shizawa 90] M. Shizawa & K. Mase. *Simultaneous multiple optical flow estimation*. In Int. Conf. on Pattern Recognition, volume 1, pages 274–278, Atlantic City, 1990.
- [Shizawa 91a] M. Shizawa & K. Mase. *Principle of superposition : A common computational framework for analysis of multiple motions*. In IEEE Workshop on Visual Motion, pages 164–172, Princeton, octobre 1991.
- [Shizawa 91b] M. Shizawa & K. Mase. *An unified computational theory for motion transparency and motion boundaries based on eigen energy analysis*. In Proc. Computer Vision and Pattern Recognition, CVPR'91, pages 289–295, Maui, juin 1991.
- [Simoncelli 92] E.P. Simoncelli. *Distributed representation of image velocity*. Rapport technique, Vision and Modeling Technical Report 202, MIT Media Laboratory, octobre 1992.
- [Spekowius 95] G. Spekowius & al. *Simulation of the imaging performance of x-ray image intensifier/TV camera chain*. SPIE, vol. 2432, pages 12–23, 1995.
- [Spinei 01] A. Spinei & D. Pellerin. *Motion estimation of opaque or transparent objects using triads of Gabor*. Signal Processing, vol. 81, no. 4, pages 845–853, avril 2001.
- [Stuke 03a] I. Stuke, T. Aach, C. Mota & E. Barth. *Estimation of multiple motions by block matching*. In 4th ACIS Int. Conf. on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD 2003), pages 358–362, Luebeck, octobre 2003.
- [Stuke 03b] I. Stuke, T. Aach, C. Mota & E. Barth. *Estimation of multiple motions : regularization and performance evaluation*. In Image and Video Communications and Processing 2003, SPIE, volume 5022, pages 75–86, Santa Clara, mai 2003.

- [Stuke 03c] I. Stuke, T. Aach, C. Mota & E. Barth. *Linear and regularized solutions for multiple motions*. In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 03), pages 157–160, Hong Kong, avril 2003.
- [Stuke 04] I. Stuke, T. Aach, C. Mota & E. Barth. *Estimation of multiple motions using block-matching and Markov random fields*. In SPIE Conf. Visual Communications and Image Processing, VCIP'2004, pages 274–278, San Jose, janvier 2004.
- [Szeliski 00] R. Szeliski, S. Avidan & P. Anandan. *Layer extraction from multiple images containing reflections and transparency*. In IEEE Conference on Computer Vision and Pattern Recognition CVPR'00, pages 246–256, Hilton Head, juin 2000.
- [Tomasi 98] C. Tomasi & R. Manduchi. *Bilateral filtering for gray and color images*. In Proc. IEEE Int. Conf. on Computer Vision, ICCV'98, pages 839–46, Bombay, Inde, 1998.
- [Toro 00a] J. Toro. *Direct and indirect parametric multiple motion estimation in image sequences and their application in layer-based angiographic image compression*. These, University of Ulster, mars 2000.
- [Toro 00b] J. Toro, F. Owens & R. Medina. *Multiple motion estimation and segmentation in transparency*. In Proc. of the IEEE Int. Conference on Acoustics, Speech and Signal Processing, pages 2087–2090, Istanbul, juin 2000.
- [Tschumperlé 02] D. Tschumperlé. *PDE's based regularization of multivalued images and applications*. These, Université de Nice-Sophia Antipolis, décembre 2002.
- [Unser 93] M. Unser, A. Aldroubi & M. Eden. *The L2 polynomial spline pyramid*. IEEE Trans. Pattern Analysis Machine Intelligence, vol. 15, no. 4, pages 364–379, 1993.
- [Uttenweiler 03] D. Uttenweiler, C. Weber, B. Jähne, R.H. Fink & H. Scharr. *Spatio-temporal anisotropic diffusion filtering to improve signal-to-noise ratios and object restoration in fluorescence microscopic image sequences*. Journal of Biomedical Optics, vol. 8, no. 1, pages 40–47, janvier 2003.
- [van Roosmalen 96] P. van Roosmalen, S.J.P. Westen, R.L. Lagendijk & J. Biemond. *Noise reduction for image sequences using an oriented pyramid thresholding technique*. In Proc. IEEE Int. Conf. on Image Processing, ICIP'96, volume 1407 of *Lecture Notes in Computer Science*, pages 375–378, Lausanne, Suisse, septembre 1996.
- [Vernon 98] D. Vernon. *Decoupling Fourier components of dynamic image sequences : A theory of signal separation, image segmentation, and optical flow estimation*. In ECCV (2), pages 69–85, Freiburg, juin 1998.
- [Weickert 99] J. Weickert. *Coherence Enhancing diffusion filtering*. International Journal of Computer Vision, vol. 31, no. 2/3, pages 111–127, 1999.

- [Weiss 01] Y. Weiss. *Deriving intrinsic images from image sequences*. In Proceedings International Conference on Computer Vision ICCV 2001, pages 68–75, Vancouver, juillet 2001.
- [Woods 93] J.W. Woods & J. Kim. Motion analysis and image sequence processing, chapitre 12, Motion compensated spatiotemporal Kalman filter. Kluwer Academic Publishers, 1993.
- [Yu 99] W. Yu, G. Sommer, S. Beauchemin & K. Daniilidis. *Detection and characterisation of multiple motion points*. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'99), pages 1171–1177, Fort Collins, USA, juin 1999.
- [Yu 03] W. Yu, G. Sommer & K. Daniilidis. *Multiple motion analysis in spatial domain or in spectral domain?* Computer Vision and Image Understanding, vol. 90, no. 2, pages 129–152, mai 2003.

Résumé : L'imagerie médicale par rayons X ouvre de nombreuses possibilités de diagnostic et de thérapie. Cependant, les rayons X étant nocifs, l'exposition doit être aussi limitée que possible, ce qui se traduit par des images bruitées et peu contrastées. Dans ce contexte, le traitement de l'image peut être d'une aide précieuse. Nous nous intéressons au problème de l'estimation des mouvements anatomiques dans les séquences d'images fluoroscopiques où le problème de la transparence des images est central. Deux hypothèses simplificatrices raisonnables, portant sur la régularité des mouvements anatomiques (menant à l'utilisation de modèles paramétriques de mouvement) et le nombre de couches transparentes pouvant se superposer en une région donnée de l'image (menant à l'introduction du concept de *bi-transparence distribuée*), permettent de formuler ce problème d'estimation de manière appréhensible. C'est sur cette base que nous proposons trois méthodes d'estimation des mouvements transparents (la plus complète d'entre elle induisant un schéma joint de segmentation et d'estimation), particulièrement robustes au bruit des images fluoroscopiques, qui permettent de traiter progressivement le problème général considéré. Leur précision est démontrée sur de nombreuses expériences sur données synthétiques réalistes et une large base de séquences d'images cliniques réelles. Nous abordons ensuite l'exploitation des mouvements estimés pour le débruitage. Le phénomène de transparence impose le développement d'une méthode spécifique de compensation des mouvements transparents. Nous étudions les propriétés des filtres avec compensation des mouvements transparents, et proposons une méthode hybride de filtrage dépassant certaines de leurs limitations.

Mots-clefs : Imagerie par rayons X, estimation de mouvements en situation de transparence, modèles paramétriques de mouvement, méthode de Gauss-Newton, champs de Markov, filtre de réduction de bruit récursif, filtrage adaptatif.

Abstract : X-ray medical imaging allows numerous diagnostic and therapeutic possibilities. However, the X-rays being noxious, their radiation has to be kept as low as possible, which results in low-contrasted noisy images. Image processing can be of great help in this context. We focus in this study on the anatomical motion estimation, which is an original topic since the processed images are transparent. Two reasonable simplifying assumptions are made, about anatomical motion regularity (leading to the use of parametric motion models) and about the number of transparent layers that can be superimposed in a given spatial region (leading to the introduction of the concept of *bi-distributed transparency*). They enable to reformulate the difficult initial problem in a tractable one. As a result, we propose three transparent motion estimation methods (the more general one involving a joint motion estimation and segmentation process) that are particularly robust to noise. They allow to progressively address the general problem of anatomical transparent motion estimation. Their accuracy is demonstrated on numerous experiments on synthetic data, as well as on a large set of clinical image sequences. In a second stage of our work, we focus on the use of the estimated motions for denoising purposes. The transparency phenomenon enforces the development of a specific transparent motion compensation procedures. We study the properties of the resulting denoising filters, and propose a hybrid method able to bypass some of their limitations.

Keywords : X-Ray imaging, motion estimation in transparency, parametric motion models, Gauss-Newton method, Markov random fields, recursive denoising filter, adaptive denoising filter.