

**Apprentissage non supervisé et suivi de modèles  
déformables dans une séquence d'images  
in image sequences**

**Charles Kervrann et Fabrice Heitz**

*IRISA/INRIA, Campus Universitaire de Beaulieu,  
35042 Rennes Cedex, France  
E-mail: kervrann@irisa.fr, heitzfa@enspsmail.u-strasbg.fr*

**10<sup>ème</sup> congrès AFCET  
– Reconnaissance des Formes et Intelligence Artificielle –  
Janvier 1996, Rennes, France**

# Apprentissage non supervisé et suivi de modèles déformables dans une séquence d'images

Charles Kervrann  
IRISA/INRIA  
Campus Universitaire de Beaulieu  
35042 Rennes Cedex, France  
E-mail: kervrann@irisa.fr

Fabrice Heitz  
ENSPS/LSIIT  
Boulevard Sébastien Brant  
67400 Illkirch, France  
E-mail: heitzfa@enspsmail.u-strasbg.fr

## Résumé

*Dans cet article, nous proposons une méthode originale (non supervisée) pour réaliser l'apprentissage et le suivi des modes de déformation d'une structure bidimensionnelle en mouvement dans une séquence d'images. La représentation de la structure d'intérêt repose sur une description hiérarchique des déformations appliquées à une forme prototype ("template"). L'estimation bayésienne des déformations est obtenue en maximisant la loi conjointe décrivant les relations entre variables cachées et observations. Une décomposition de Karhunen-Loeve des déformations estimées permet d'extraire les modes de déformation dominants observés dans la séquence. Cette procédure récursive, ne requérant que la définition d'une forme prototype initiale grossière, est complétée par un suivi temporel des paramètres de déformation déjà identifiés. Des expérimentations variées sur des études de cas réelles illustrent la validité de cette approche pour le suivi de structures déformables.*

**Mots-Clefs :** modèles déformables, Karhunen-Loeve, segmentation, mouvement, suivi, filtrage de Kalman.

## 1 Introduction

La modélisation des déformations et leur exploitation en segmentation de séquences d'images ont connu un développement important ces cinq dernières années, développement dicté en particulier par les besoins croissants des applications dans ce domaine (imagerie biomédicale - analyse du mouvement humain - mécanique des fluides). L'introduction de "modèles déformables" [12, 25, 21, 23] pour la représentation des déformations s'est révélée importante à plus d'un titre :

- d'une part, ces modèles permettent la prise en compte d'informations *a priori* sur la structure des objets, ce qui introduit une robustesse appréciable dans les techniques d'extraction ou de suivi d'objets déformables ;
- d'autre part, la modélisation des déformations ouvre des perspectives prometteuses, mais peu explorées

pour l'instant, dans le domaine de la caractérisation et de l'interprétation du comportement dynamique d'objets complexes.

D'abord utilisés en segmentation d'images statiques, ces modèles sont désormais exploités avec profit pour le suivi de structures déformables dans des séquences d'images longues ou pour la caractérisation d'objets par l'analyse des déformations [6, 10, 19]. Par ailleurs, les modèles que nous présentons dans cet article s'appliquent essentiellement aux classes d'objets polyarticulés et aux classes d'objets structurés présentant des déformations élastiques.

Des travaux antérieurs sur les contours actifs et leurs déclinaisons ont permis d'intégrer des contraintes génériques sur la courbe plane modélisant le contour d'une structure déformable [12, 3, 22]. Ces modèles génériques déterministes ne sont cependant pas toujours bien adaptés aux contraintes spécifiques qu'il est souhaitable d'imposer dans une application particulière [3]. Des modèles paramétriques, offrant une description plus spécifique de la forme à représenter ont également été proposés [26, 5]. Ces modèles sont construits "à la main" à partir de courbes ou surfaces paramétriques simples (cercles, coniques, hyperquadriques, etc.) liées entre elles par des contraintes.

Dans [13] nous avons introduit un nouveau cadre de modélisation (statistique) permettant d'intégrer des connaissances *a priori* dans la segmentation et le suivi de structures déformables bidimensionnelles en mouvement. La représentation de la forme s'appuie sur une description statistique hiérarchique des déformations appliquées à une forme prototype ("template"). Dans [13] les modes de déformations globaux de la structure 2-D sont identifiés au moyen d'une décomposition de Karhunen-Loeve (KL) des déformations observées sur une population d'apprentissage. La représentation a donc pour caractéristique de s'adapter (d'un point de vue statistique) à la structure et aux déformations particulières observées. Une configuration de l'objet se déduit de la forme prototype ("template") par des transformations géométriques globales (similitude plane) et des transformations ren-

dant compte des modes globaux de déformation les plus significatifs. Des déformations locales, permettant de raffiner la solution, sont également introduites et sont modélisées par des processus markoviens [10, 13, 16]. Ce modèle hiérarchique est ensuite intégré comme modèle *a priori* dans un schéma d'estimation bayésienne pour extraire la structure d'intérêt de la séquence d'images [13].

L'approche développée dans [13] nécessite cependant une phase d'apprentissage (réalisée de façon manuelle en mettant en correspondance des formes dans une séquence d'images représentatives). Nous proposons dans cet article une nouvelle méthode d'apprentissage *non supervisée* des modes de déformation, plus robuste aux occlusions partielles et aux problèmes de mises en correspondance et qui ne requiert que la définition d'une forme prototype initiale grossière de l'objet d'intérêt. Dans cette nouvelle approche la structure de l'objet est remise à jour au cours du temps et les modes de déformation sont réactualisés au moyen d'une décomposition de Karhunen Loeve des formes précédemment estimées. Une représentation stationnaire des déformations et de la structure est obtenue en pratique lorsqu'un ensemble suffisant de configurations représentatives de l'objet d'intérêt a été observé. La prise en compte de la cohérence du mouvement de la structure déformable améliore par ailleurs sensiblement la qualité de la segmentation. Dans ce but, un suivi des paramètres de déformation par filtrage récursif [14] est intégré dans la boucle de segmentation. Le suivi par filtrage de Kalman [1, 2, 4, 8, 17, 21] permet de prendre en compte des déplacements importants de l'objet d'intérêt, facilitant l'optimisation lors de l'estimation des paramètres de déformation [14].

L'article est organisé de la façon suivante. Nous rappelons tout d'abord en section 2, les caractéristiques du modèle déformable hiérarchique adopté dans notre approche. La procédure de segmentation et les techniques d'optimisation qui lui sont associées sont rappelées en section 3. Nous décrivons ensuite (en section 4) la méthode non supervisée d'apprentissage automatique de la structure et des modes de déformation. La technique de suivi temporel du modèle est présentée dans la section 5<sup>1</sup>. Des exemples d'apprentissage, de segmentation et de suivi de structures déformables variées sur des séquences réelles, sont proposés en section 4 et 5. Ils illustrent l'aptitude de l'approche proposée à s'adapter à des applications diverses ainsi que sa robustesse dans le traitement de séquences longues (plus d'une centaine de trames).

## 2 Un modèle déformable statistique hiérarchique

Les modèles déformables introduits en imagerie intègrent des connaissances *a priori* sur la structure des

1. Une version préliminaire de cet article, avec un nombre très limité de résultats expérimentaux a été présenté dans [15].

objets et sur les déformations de ces structures. Dans ce contexte, les techniques de décomposition modale (issues de la mécanique) se sont révélées très efficaces pour décrire le comportement dynamique d'objets complexes naturels [24, 21, 19, 20]. Une autre approche consiste à décomposer les déformations sur une base particulière; une base de Fourier a par exemple été retenue par Staib et Duncan [23] pour segmenter des objets déformables en imagerie médicale. Dans le cadre de notre travail, la représentation des déformations globales est obtenue grâce à une technique récemment décrite par Cootes *et al.* [6] qui proposent de retenir un nombre réduit de modes significatifs de variation dans une décomposition de Karhunen-Loeve des déformations observées sur une population représentative. La mise en œuvre du modèle déformable passe par une phase préalable d'apprentissage (hors-ligne). Une forme particulière  $\mathbf{x}$  est représentée par le vecteur  $(2n \times 1)$  des  $n$  points caractéristiques de son contour, définissant une approximation polygonale de l'objet à représenter. Les déformations d'une forme  $\mathbf{x}$ , observées sur une population représentative, sont ensuite caractérisées par un vecteur de déplacement  $d\mathbf{x} = \mathbf{x} - \bar{\mathbf{x}}$  où  $\bar{\mathbf{x}}$  représente le modèle moyen ("template").

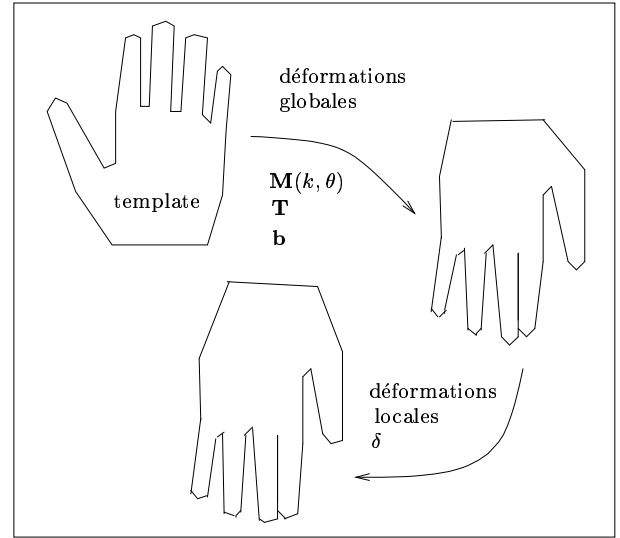


FIG. 1 - Description hiérarchique des déformations.

La décomposition de Karhunen-Loeve s'obtient par diagonalisation de la matrice d'autocovariance du vecteur de déplacement  $d\mathbf{x}$ . Elle conduit à des paramètres de déformation qui sont décorrélés et autorise une représentation compacte en ne conservant que les modes de variation  $\mathbf{b}$  les plus significatifs. Le changement de base, associé à cette transformation de Karhunen-Loeve s'écrit :

$$\mathbf{b} = \Phi^T d\mathbf{x} = \Phi^T (\mathbf{x} - \bar{\mathbf{x}}) \quad (1)$$

où  $\Phi$  représente la matrice  $(2n \times m)$  des  $m$  vecteurs propres unitaires correspondant aux  $m$  valeurs propres les plus fortes,  $\mathbf{b}$  désigne le vecteur  $(m \times 1)$  correspondant aux  $m$  modes de déformation les plus significatifs

exprimés dans la nouvelle base. Le modèle déformable est alors représenté comme suit [13] :

$$\mathbf{X} = \mathbf{M}(k, \theta) [\bar{\mathbf{x}} + \Phi \mathbf{b}] + \mathbf{T} \quad (2)$$

où le vecteur de translation  $\mathbf{T}$ , la rotation d'angle  $\theta$  et l'homothétie de facteur  $k$  (regroupées dans la matrice de transformation  $\mathbf{M}(k, \theta)$ ) ainsi que les modes de déformation globaux  $\mathbf{b}$  définissent les transformations géométriques globales (similitudes planes+déformations), appliquées à la forme prototype.

Un processus de déformation locale  $\delta$ , appliqué aux  $n$  points, est introduit pour affiner cette première description. Les déformations locales peuvent être vues comme des perturbations aléatoires se superposant à la structure globalement déformée. Une hypothèse raisonnable consiste à supposer que  $\delta$  suit un modèle markovien-gaussien, modélisant les interactions entre points voisins [13]. Le modèle déformable complet peut alors s'écrire sous la forme (fig. 1) :

$$\mathbf{Y} = \mathbf{M}(k, \theta) [\bar{\mathbf{x}} + \Phi \mathbf{b}] + \mathbf{T} + \delta. \quad (3)$$

Dans notre cas,  $\delta = (\delta_1, \delta_2, \dots, \delta_n)$  et  $\delta_i = (\delta_{x_i}, \delta_{y_i})$ , est un processus markovien-gaussien centré du 1<sup>er</sup> ordre :

$$\mathbf{P}(\delta) = \frac{1}{Z_p} \exp -\frac{1}{2} \delta^T \mathbf{R}^{-1} \delta \quad (4)$$

où  $\mathbf{R}$  est la matrice de covariance de  $\delta$  et  $Z_p$  est une constante de normalisation. Sous cette hypothèse, la loi conjointe de  $\delta$  s'écrit :

$$\mathbf{P}(\delta) = \frac{1}{Z_p} \exp -\frac{1}{2} \sum_{i=1}^n \left[ \frac{1}{\epsilon_i^2} \|\delta_i - \delta_{i-1}\|^2 + \frac{1}{\sigma_i^2} \|\delta_i\|^2 \right] \quad (5)$$

où  $\sigma_i^2$  et  $\epsilon_i^2$  sont les paramètres de variance du modèle ;  $\epsilon_i^2$  agit sur les interactions entre points voisins tandis que des valeurs faibles pour  $\sigma_i^2$  ont tendance à minimiser l'effet des déformations locales.

### 3 Un cadre statistique pour la segmentation

Nous rappelons ici les principes généraux de la technique statistique de segmentation développée dans [13]. Elle repose sur une formulation bayésienne du problème, intégrant le modèle *a priori* défini ci-dessus pour la structure déformable ainsi que des modèles d'observations dépendants de l'application traitée.

#### 3.1 Estimation bayésienne des déformations

Nous nous intéressons au problème particulier de l'extraction d'objets mobiles (déformables) d'une séquence d'images, dans des situations où la caméra est fixe. Soit  $\mathbf{d} = (d_s, s \in S)$  l'ensemble des observations définies sur une grille rectangulaire (les observations sont liées aux gradients spatio-temporels extraits de la séquence

d'images). Dans ce cadre, le problème de la segmentation est posé comme étant celui de l'estimation (au sens du Maximum A Posteriori – MAP) de la réalisation (cachée) du processus  $\mathbf{Y}$  portant la structure déformable à partir du champ des observations  $\mathbf{d}$  :

$$\mathbf{Y}^* = \arg \max_{\mathbf{Y}} \mathbf{P}(\mathbf{d} | \mathbf{Y}) \mathbf{P}(\mathbf{Y}). \quad (6)$$

D'après les hypothèses statistiques énoncées plus haut (équ. 4),  $\mathbf{Y}$  est un processus markovien-gaussien du 1<sup>er</sup> ordre :

$$\mathbf{P}(\mathbf{Y}) = \frac{1}{Z_p} \exp -\frac{1}{2} (\mathbf{Y} - \mathbf{X})^T \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X}). \quad (7)$$

Les paramètres de similitude plane  $\mathbf{M}(k, \theta)$ ,  $\mathbf{T}$  et les modes de déformation globaux  $\mathbf{b}$  sont ici considérés comme des paramètres déterministes de ce modèle probabiliste. On notera  $\Theta = (\mathbf{M}(k, \theta), \mathbf{T}, \mathbf{b})$  les *hyperparamètres* du modèle  $\mathbf{Y}$ .

La loi  $\mathbf{P}(\mathbf{d} | \mathbf{Y})$  exprime la vraisemblance des observations conditionnellement au processus de déformation à estimer ; elle est spécifiée de façon "ad hoc" par une distribution de Gibbs [9] qui intègre des connaissances spécifiques liées au type d'images ou d'objets à traiter :

$$\mathbf{P}(\mathbf{d} | \mathbf{Y}) = \frac{1}{Z_d} \exp -E_d(\mathbf{Y}, \mathbf{d}) \quad (8)$$

où  $E_d(\mathbf{Y}, \mathbf{d})$  est une fonction d'énergie et  $Z_d$  est une constante de normalisation. Si  $\Theta$  représente le vecteur des hyperparamètres, l'estimée au sens du MAP des déformations est définie par :

$$\mathbf{Y}^* = \arg \max_{\mathbf{Y}} \mathbf{P}(\mathbf{d} | \mathbf{Y}) \mathbf{P}_{\Theta}(\mathbf{Y}). \quad (9)$$

La distribution conjointe est alors également une distribution de Gibbs :

$$\mathbf{P}(\mathbf{d} | \mathbf{Y}) \mathbf{P}_{\Theta}(\mathbf{Y}) = \frac{1}{Z} \exp -E_{\Theta}(\mathbf{Y}, \mathbf{d}), \quad (10)$$

dont la constante de normalisation  $Z$  ne dépend pas de  $\Theta$ . Estimer l'ensemble du processus de déformation au sens du MAP est équivalent au problème de la minimisation d'une fonction d'énergie  $E_{\Theta}(\mathbf{Y}, \mathbf{d})$  par rapport aux paramètres du modèle :

$$E_{\Theta}(\mathbf{Y}, \mathbf{d}) = E_p(\mathbf{Y}) + E_d(\mathbf{Y}, \mathbf{d}). \quad (11)$$

Soulignons que la modélisation du seul terme de vraisemblance  $E_d(\mathbf{Y}, \mathbf{d})$  conduirait à des résultats de segmentation peu robustes au bruit et à la variabilité des objets observés. L'introduction du modèle *a priori*  $E_p(\mathbf{Y})$  permet de contourner cette difficulté en *contraignant* le processus de déformation. Toute la difficulté du problème réside à présent dans l'estimation conjointe du champ  $\mathbf{Y}$  et du vecteur des hyperparamètres  $\Theta$ .

#### 3.2 Maximum de Vraisemblance Marginale

De manière générale, la loi conjointe (équ. 10) est une fonction non-linéaire des paramètres du modèle  $\mathbf{M}(k, \theta)$ ,

**T** et **b**. Par commodité d'écriture, le modèle stochastique est redéfini comme :

$$\mathbf{Y} = \mathbf{X}(\Theta) + \delta. \quad (12)$$

Puisque  $\Theta$  est inconnu, le problème de segmentation revient à estimer le couple  $(\Theta^*, \mathbf{Y}^*)$  selon le critère d'optimalité :

$$(\Theta^*, \mathbf{Y}^*) = \arg \max_{\Theta, \mathbf{Y}} \mathbf{P}(\mathbf{Y}, \mathbf{d} | \Theta). \quad (13)$$

Ce critère satisfait le maximum global, par rapport à  $\mathbf{Y}$  et  $\Theta$ , de la distribution  $\mathbf{P}(\mathbf{Y}, \mathbf{d} | \Theta)$  mais est extrêmement difficile à implanter sous cette forme. Nous sommes tout à fait libre de ne pas apprécier  $\mathbf{Y}$  et  $\Theta$  comme des processus se situant au même niveau ; c'est effectivement le cas pour notre problématique [18]. Dans ces conditions, nous considérons l'estimateur du *Maximum de Vraisemblance Marginale* (MVM) pour déterminer  $\Theta$  [18] :

$$\Theta^* = \arg \max_{\Theta} \int_{\mathbf{Y}} \mathbf{P}(\mathbf{Y}, \mathbf{d} | \Theta) d\mathbf{Y}. \quad (14)$$

Cette estimée est exploitée dans une seconde étape pour l'estimation de la solution au problème considéré :

$$\mathbf{Y}^* = \arg \max_{\mathbf{Y}} \mathbf{P}(\mathbf{Y} | \mathbf{d}, \Theta^*). \quad (15)$$

**Estimation de  $\Theta$**  Nous proposons ici un estimateur de  $\Theta$  permettant d'approcher l'estimateur MVM (équ. 14). D'après la règle de Bayes, nous pouvons écrire :

$$\mathbf{P}(\mathbf{Y}, \mathbf{d} | \Theta) = \mathbf{P}(\mathbf{d} | \mathbf{Y}, \Theta) \mathbf{P}(\mathbf{Y} | \Theta). \quad (16)$$

Une hypothèse raisonnable consiste à supposer que le processus aléatoire  $\delta$  est de variance faible ( $\sigma_i^2 \rightarrow 0, \forall i$ ).  $\mathbf{Y}$  reste très "concentré" autour de  $\mathbf{X}(\Theta)$ . Dans ces conditions, la loi gaussienne de  $\mathbf{Y}$  peut être approchée par une masse de Dirac :

$$\mathbf{P}(\mathbf{Y} | \Theta) \approx \delta(\mathbf{Y} - \mathbf{X}(\Theta)) \quad (17)$$

où  $\delta(\cdot)$  désigne la masse de Dirac. Il vient alors :

$$\begin{aligned} \Theta^* &= \arg \max_{\Theta} \int_{\mathbf{Y}} \mathbf{P}(\mathbf{d} | \mathbf{Y}, \Theta) \mathbf{P}(\mathbf{Y} | \Theta) d\mathbf{Y} \\ &= \arg \max_{\Theta} \int_{\mathbf{Y}} \mathbf{P}(\mathbf{d} | \mathbf{Y}, \Theta) \delta(\mathbf{Y} - \mathbf{X}(\Theta)) d\mathbf{Y} \\ &= \arg \max_{\Theta} \mathbf{P}(\mathbf{d} | \mathbf{Y} = \mathbf{X}(\Theta)). \end{aligned} \quad (18)$$

Nous proposons enfin un estimateur de  $\Theta$  vérifiant :

$$\Theta^* = \arg \min_{\Theta} E_d(\mathbf{X}(\Theta), \mathbf{d}). \quad (19)$$

**Estimation de  $\mathbf{Y}$**  Une solution pour estimer  $\mathbf{Y}$  au sens du MAP est de vérifier le critère déjà proposé dans (équ. 15) connaissant l'estimée optimale  $\Theta^*$ . D'après (équ.10), il est clair que ce critère est équivalent à minimiser une fonction d'énergie globale :

$$\mathbf{Y}^* = \arg \min_{\mathbf{Y}} E_{\Theta^*}(\mathbf{Y}, \mathbf{d}). \quad (20)$$

### 3.3 Modèles d'observation pour la segmentation

Pour la classe des problèmes qui nous intéressent ici, une énergie d'interaction entre processus de déformation et observations  $E_d(\mathbf{Y}, \mathbf{d})$ , fonction des paramètres du modèle déformable, est définie. La fonction d'énergie  $E_d(\mathbf{Y}, \mathbf{d})$  que nous spécifions s'appuie sur des observations liées aux gradients spatio-temporels extraits de la séquence d'image. Elle est bâtie différemment selon la nature des observations disponibles : extraction de régions déformables homogènes au sens du mouvement ou extraction des contours photométriques d'un objet déformable (nous proposons dans la suite de cet article plusieurs exemples venant illustrer ces deux cas).

Soit  $I_s^t$ ,  $s \in S$  l'image de luminance à l'instant  $t$  où  $s = (x, y)$  désigne les coordonnées spatiales. Dans le contexte de la détection d'objets déformables en mouvement, les changements temporels seuillés de la fonction luminance sont évalués en chaque site  $s$  de l'image. La distribution  $\mathbf{P}(\mathbf{d} | \mathbf{Y})$  décrit dans ce cas l'interaction entre les variables observables  $\mathbf{d}$  et la configuration du modèle déformable. Pour une configuration donnée du modèle déformable, l'image est partitionnée en deux régions : l'intérieur du modèle  $\Gamma_{\mathbf{Y}}^I$  et l'extérieur du modèle  $\Gamma_{\mathbf{Y}}^O$  correspondant au fond. La distribution définie dans ce cas vise à regrouper à l'intérieur du modèle déformable les points de l'image présentant un mouvement significatif et à rejeter à l'extérieur du modèle les points appartenant au fond (fixe) de l'image :

$$\mathbf{P}(\mathbf{d} | \mathbf{Y}) = \frac{1}{Z_d} \exp - \left\{ \sum_{s \in \Gamma_{\mathbf{Y}}^I} |d_s - 1| + \sum_{s \in \Gamma_{\mathbf{Y}}^O} |d_s - 0| \right\} \quad (21)$$

où  $Z_d$  est une constante de normalisation.

Pour la résolution du problème (équ. 19 et équ. 20), les algorithmes de minimisation classiques (du type gradient) présentent l'inconvénient de se piéger dans des minima locaux, pouvant correspondre à des solutions de qualité médiocre. Les techniques stochastiques, peu sensibles à l'initialisation mais plus coûteuses, ne présentent pas cet inconvénient. L'optimisation globale utilisée dans [13] s'appuie sur un algorithme de recuit simulé basé sur l'échantillonneur de Gibbs [9] et peut être mise en œuvre dans des conditions calculatoires acceptables puisque le nombre de paramètres à optimiser est réduit.

Par ailleurs, dans de nombreuses applications, les contours photométriques (gradients spatiaux) de l'objet d'intérêt représentent la seule information exploitable pour réaliser la segmentation. Ces observations permettent cependant une localisation précise de l'objet, sous réserve que le modèle déformable soit convenablement initialisé, manuellement ou de manière automatique, par des procédures adéquates (morphologie mathématique [22], transformée de Hough [16]). Dans ce cas, l'énergie

globale est spécifiée par :

$$E_{\Theta}(\mathbf{Y}, \mathbf{d}) = \sum_{s \in \Gamma_{\mathbf{Y}}} \|\vec{\nabla} I_s\| + \frac{1}{2}(\mathbf{Y} - \mathbf{X})^T \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X}) \quad (22)$$

où  $\|\vec{\nabla} I_s\|$  représente la norme du gradient au site  $s$  et  $\Gamma_{\mathbf{Y}}$  désigne la frontière de la structure déformable. L'énergie  $E_d(\mathbf{Y}, \mathbf{d})$  est ici définie comme l'intégrale de la norme du gradient observée le long de la frontière définie par le modèle. On adopte dans ce cas un schéma d'optimisation similaire au cas précédent. Soulignons toutefois qu'une optimisation globale par des techniques stochastiques n'est pas appropriée ici : le critère de segmentation n'est valide que dans une *région d'intérêt* dans le voisinage de la solution désirée. Une optimisation par descente de gradient, s'appuyant sur le calcul analytique des dérivés partielles de la fonction d'énergie, est utilisée dans ce contexte.

Dans la suite de l'article, on admet que le modèle est convenablement initialisé dans la première image soit par des techniques stochastiques [13] (détection d'objets déformables en mouvement) ou de manière manuelle (extraction des contours photométriques de l'objet) [14].

## 4 Apprentissage non supervisé des modes de déformation

Dans cette section, nous abordons le problème de l'apprentissage non-supervisé de la structure et des modes de déformation du modèle défini en section 2. L'approche développée par Cootes *et al.* [7] et reprise dans [13] nécessite une phase d'apprentissage hors ligne réalisée de façon manuelle. Cette tâche, qui peut rapidement s'avérer fastidieuse, est réalisée par mise en correspondance point à point des formes appartenant à l'ensemble d'apprentissage. Des problèmes peuvent alors se poser lorsque la séquence d'apprentissage contient des situations d'occlusions partielles ou totales.

Dans cette section, nous décrivons une nouvelle méthode d'apprentissage non-supervisé (en ligne) de la structure et des modes de déformation au cours du temps, ne nécessitant que la définition d'une forme prototype initiale approchée.

### 4.1 Description générale de la méthode

Dans le cadre d'un apprentissage en ligne des caractéristiques du modèle déformable, la représentation est redéfinie à l'instant  $t$  par :

$$\mathbf{Y}_t = \mathbf{M}(k_t, \theta_t) [\bar{\mathbf{x}}_t + \Phi_t \mathbf{b}_t] + \mathbf{T}_t + \delta_t. \quad (23)$$

Le modèle moyen à l'instant  $t$  est noté  $\bar{\mathbf{x}}_t$  et  $\Phi_t$  désigne la matrice des vecteurs propres unitaires associés aux  $m_t$  valeurs propres les plus importantes de la transformée (KL) à l'instant  $t$ . Ces différents paramètres structurels du modèle déformable dépendent maintenant du temps. L'apprentissage en ligne des caractéristiques du modèle

induit une modification du nombre  $m_t$  de modes de déformation  $\mathbf{b}_t$  au cours de la séquence analysée. L'enchaînement des différentes phases de l'apprentissage non supervisé au cours du temps se présente de la manière suivante :

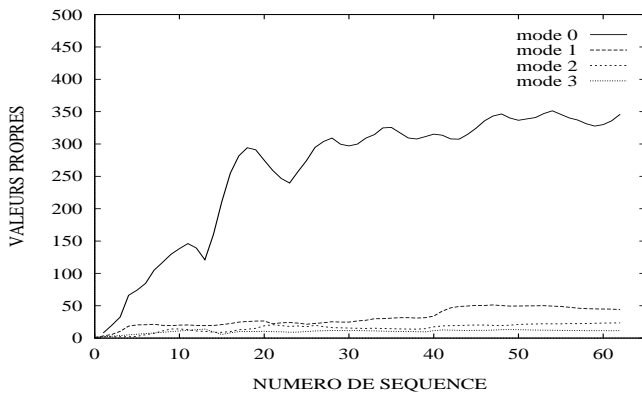
- **Traitement de la 1<sup>ère</sup> image ( $t = 1$ ) :**
  - estimation des paramètres  $\mathbf{M}(k_1, \theta_1)$  et  $\mathbf{T}_1$
  - estimation du processus de déformation locale  $\delta_1$
- **Traitement de la 2<sup>ème</sup> image ( $t = 2$ ) :**
  - estimation des paramètres  $\mathbf{M}(k_2, \theta_2)$  et  $\mathbf{T}_2$
  - estimation du processus de déformation locale  $\delta_2$
  - décomposition KL des formes déjà estimées  $\mathbf{Y}_t$  ( $t = 1, 2$ ) et détermination de  $\bar{\mathbf{x}}_2$  et  $\Phi_2$
- ⋮
- **Traitement de la  $i^{\text{ème}}$  image ( $t = i$ ) :**
  - estimation des paramètres  $\mathbf{M}(k_i, \theta_i)$  et  $\mathbf{T}_i$
  - estimation des paramètres de déformation  $\mathbf{b}_i$
  - estimation du processus de déformation locale  $\delta_i$
  - décomposition KL des formes déjà estimées  $\mathbf{Y}_t$  ( $t = 1, \dots, i$ ) et remise à jour de  $\bar{\mathbf{x}}_i$  et  $\Phi_i$ .
  - projection des modes de déformation  $\mathbf{b}_i$  dans la nouvelle base pour initialisation de  $\mathbf{b}_{i+1}$ .

Dans cette méthode d'apprentissage en ligne, les hyperparamètres  $\mathbf{M}(k_t, \theta_t)$ ,  $\mathbf{T}_t$  et  $\mathbf{b}_t$  sont estimés selon le critère du MVM et  $\mathbf{Y}_t$  est estimé au sens du MAP (section 3.2). Le traitement de la première image est spécifique, dans la mesure où l'on dispose, à cet instant, de très peu de connaissances *a priori* sur la structure (initialisation par une forme approchée) et aucune information sur les modes de déformation. L'estimation de  $\delta_1$  est donc soit menée par une technique d'optimisation stochastique (coûteuse d'un point de vue calculatoire), soit assurée par une technique de gradient, à partir d'une configuration initiale proche de la solution (obtenue par initialisation manuelle). Soulignons que le surcoût induit par l'utilisation d'un algorithme stochastique pour l'estimation de  $\delta_t$  ne concerne que les premières trames (un algorithme déterministe est ensuite utilisé systématiquement).

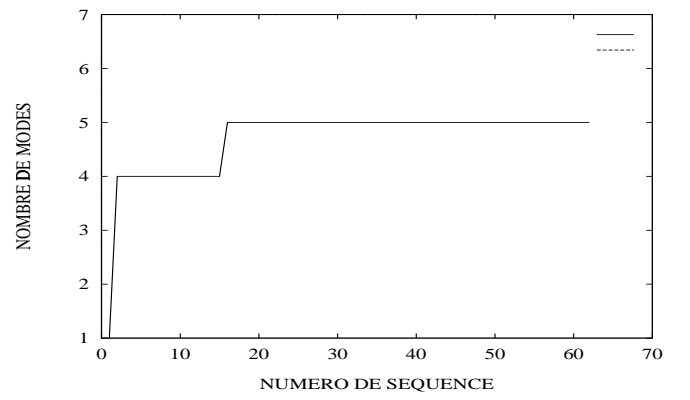
La cohérence temporelle des déplacements dans la séquence est par ailleurs exploitée en adoptant pour valeurs initiales des paramètres de déformations  $\mathbf{b}_{t+1}$  dans l'image à l'instant  $t+1$ , les valeurs finales des paramètres obtenus sur l'image à l'instant  $t$ , (exprimés dans la nouvelle base). Les formules et procédures de remise à jour sont données dans la section suivante.

### 4.2 Formules de remise à jour des déformations

Pour la remise à jour des paramètres structurels du modèle (forme moyenne  $\bar{\mathbf{x}}_t$  et vecteurs propres de la décomposition modale  $\Phi_t$ ), les formes préalablement estimées  $\mathbf{Y}_t$  dans la séquence sont analysées dans le même



(a) Evolution temporelle des valeurs propres ;



(b) Evolution temporelle du nombre de modes ;

FIG. 2 - Résultats de l'apprentissage non supervisé.

repère, en compensant, à chaque instant les transformations géométriques liées à la similitude plane. La structure déformable, s'écrit dans ce repère commun :

$$\begin{aligned} \mathbf{y}_t &= \mathbf{M}^{-1}(k_t, \theta_t) [\mathbf{Y}_t - \mathbf{T}_t], \\ \mathbf{y}_t &= \bar{\mathbf{x}}_t + \Phi_t \mathbf{b}_t + \mathbf{M}^{-1}(k_t, \theta_t) \delta_t. \end{aligned} \quad (24)$$

Le modèle moyen ("template") est donc remis à jour à l'instant  $t + 1$  selon :

$$\bar{\mathbf{x}}_{t+1} = \frac{N_t}{N_t + 1} \bar{\mathbf{x}}_t + \frac{1}{N_t + 1} \mathbf{y}_t, \quad (25)$$

$$\bar{\mathbf{x}}_{t+1} = \bar{\mathbf{x}}_t + \frac{1}{N_t + 1} [\Phi_t \mathbf{b}_t + \mathbf{M}^{-1}(k_t, \theta_t) \delta_t] \quad (26)$$

où  $N_t$  représente le nombre d'estimées  $\mathbf{Y}_t$  obtenues depuis le début de la séquence. Les modes déformations sont décrit par les vecteurs propres unitaires  $\Phi_{t+1}$  de la matrice d'autocovariance définie à l'instant  $t + 1$  par :

$$\mathbf{C}_{t+1} = \frac{1}{N_t + 1} \sum_{i=1}^{t+1} (\mathbf{y}_i - \bar{\mathbf{x}}_{t+1})^T (\mathbf{y}_i - \bar{\mathbf{x}}_{t+1}). \quad (27)$$

Pour obtenir une description compacte, une troncature des modes de déformation est réalisée en ne retenant que les  $m_{t+1}$  vecteurs propres associées aux  $m_{t+1}$  valeurs propres les plus importantes. Le nombre de modes  $m_t$  est ajusté au cours du temps de manière à conserver entre 99% et 99.5% de la variabilité totale observée (fig. 2b).

D'après les équations 26 et 27, on voit que, sous des hypothèses peu restrictives,  $\bar{\mathbf{x}}_{t+1}$  et  $\mathbf{C}_{t+1}$  convergent lorsque  $t \rightarrow \infty$ . En pratique, la convergence est atteinte lorsque l'ensemble des configurations représentatives de l'objet d'intérêt a été observé. Ceci se produit par exemple après une cinquantaine de trames, dans l'exemple de la figure 2a qui présente une étude expérimentale portant sur l'évolution des valeurs propres, dans le contexte de l'apprentissage des paramètres d'une main en mouvement.

Enfin, pour exploiter la cohérence temporelle du mouvement, l'estimée initiale des modes de déformation  $\hat{\mathbf{b}}_{t+1}$

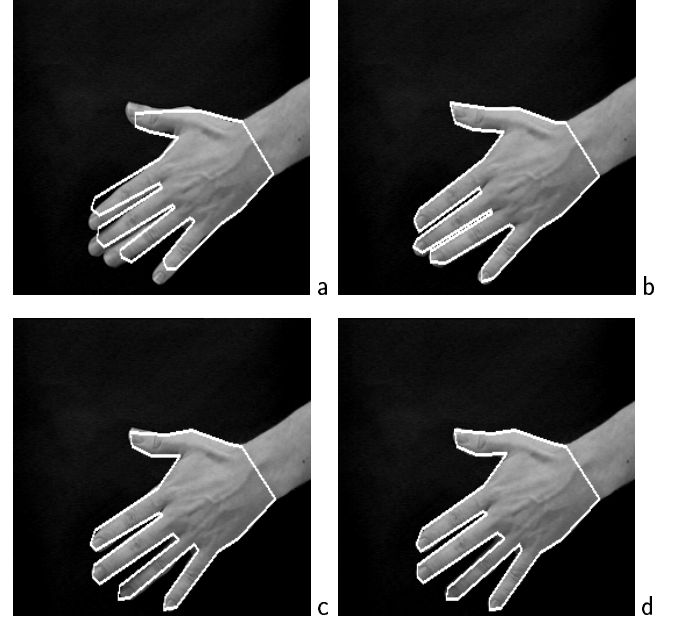


FIG. 3 - Apprentissage des modes déformation d'une main en mouvement sur un fond uniforme.

à l'instant  $t + 1$  est le résultat de la projection de l'estimée finale  $\mathbf{b}_t$ , exprimée dans la nouvelle base :

$$\hat{\mathbf{b}}_{t+1} = \Phi_{t+1}^T (\mathbf{y}_t - \bar{\mathbf{x}}_{t+1}). \quad (28)$$

i.e. :

$$\hat{\mathbf{b}}_{t+1} = \Phi_{t+1}^T (\bar{\mathbf{x}}_t + \Phi_t \mathbf{b}_t + \mathbf{M}^{-1}(k_t, \theta_t) \delta_t - \bar{\mathbf{x}}_{t+1}). \quad (29)$$

Dans ces conditions, si  $\|\Phi_{t+1}^T - \Phi_t^T\| \simeq 0$  et  $\|\bar{\mathbf{x}}_{t+1} - \bar{\mathbf{x}}_t\| \simeq 0$  lorsque  $t \rightarrow \infty$  alors  $\hat{\mathbf{b}}_{t+1}$  ne dépend plus du modèle moyen  $\bar{\mathbf{x}}_t$  :

$$\hat{\mathbf{b}}_{t+1} = \mathbf{b}_t + \Phi_t^T \mathbf{M}^{-1}(k_t, \theta_t) \delta_t. \quad (30)$$

### 4.3 Résultats expérimentaux

L'étude illustrant cette approche concerne l'apprentissage de la structure et des modes de déformation d'une

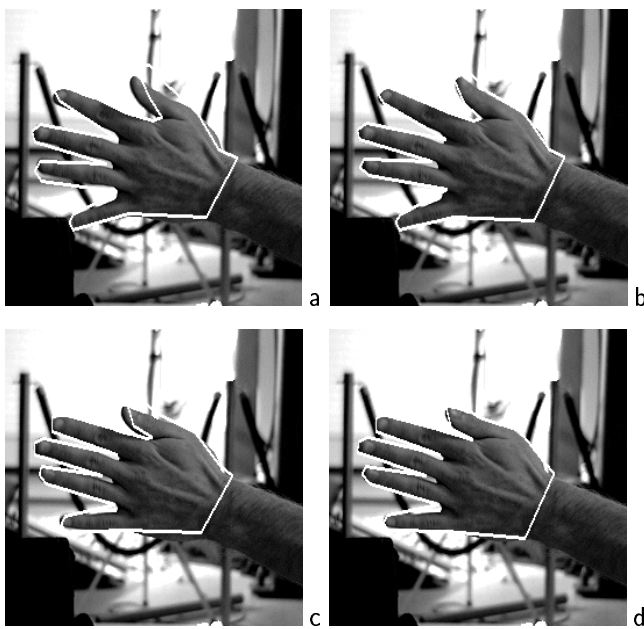


FIG. 4 - Apprentissage des modes de déformation d'une main en mouvement sur un fond texturé.

main en mouvement (fig. 3 et 4). Les observations considérées dans ce cas sont des cartes de gradients temporels seuillés. La figure 3a présente un résultat intermédiaire, correspondant à l'estimation des paramètres de similitude plane sur la 2<sup>ème</sup> image ( $t = 2$ ). A cet instant, les modes de déformation n'ont pas encore été identifiés et ne peuvent donc intervenir dans l'estimation de la forme. Seul est exploité ici le processus de déformation locale  $\delta_2$  qui est estimé grâce à un algorithme stochastique et permet de corriger le résultat obtenu par simple prise en compte des similitudes planes (résultat présenté en 3b). A l'issue de l'instant  $t = 2$ , un premier apprentissage des vecteurs propres de déformation  $\Phi_2$ , supports des modes de déformation est mené, à partir des formes obtenues aux instants  $t = 1$  et  $t = 2$ . Les modes de déformation interviennent alors dans l'estimée à l'instant  $t = 3$ . La figure 3c présente le résultat intermédiaire de l'estimation, à l'instant  $t = 3$ , des paramètres de similitude planes ainsi que des modes de déformation  $\mathbf{b}_3$ . On voit que ce résultat est très proche de la solution finale, obtenue en raffinant cette configuration intermédiaire grâce à l'optimisation du processus de déformation locale  $\delta_3$  (fig. 3d).

On constate ainsi expérimentalement que le processus de déformation locale contribue de moins en moins à l'estimation des déformations, dans la mesure où ces déformations sont représentées de façon de plus en plus précise par leurs modes globaux  $\mathbf{b}_i$ .

La figure 4 décrit les mêmes étapes du processus d'identification sur une autre séquence, présentant un fond texturé, pour deux instants consécutifs ( $t \geq 3$ ). L'apprentissage du mouvement cardiaque en angiographie a également été envisagé avec cette méthode.

## 5 Suivi du modèle déformable statistique

La prédiction heuristique proposée dans la section 4.2, (équ. 28) peut être remplacée avec profit par un suivi temporel des hyperparamètres du modèle, utilisant un filtrage récursif (filtre de Kalman) que nous décrivons maintenant. Un suivi des paramètres du modèle déformable permet de tenir compte des déplacements importants de l'objet d'intérêt dans la séquence, garantissant des segmentations robustes et permettant de faciliter l'optimisation des hyperparamètres du modèle par une meilleure initialisation d'un instant à l'autre [4, 2].

### 5.1 Suivi des hyperparamètres du modèle par filtrage de Kalman

On construit un estimateur récursif linéaire des hyperparamètres  $\Theta$  du modèle. Le processus de déformation locale n'est pas pris en compte dans les équations du filtre car trop sensible au bruit inhérent aux observations. Les mesures considérées dans ce filtre, sont le résultat de l'estimation, au sens du sc mvm des hyperparamètres. Le suivi n'est possible que lorsque  $\|\Phi_{t+1}^T - \Phi_t^T\| \simeq 0$  et  $\|\bar{\mathbf{x}}_{t+1} - \bar{\mathbf{x}}_t\| \simeq 0$ : les statistiques sur les déformations, obtenues manuellement ou automatiquement (section 4), sont alors invariantes dans le temps.

L'évolution temporelle du modèle s'exprime au moyen d'un développement du 2<sup>ème</sup> ordre de  $\mathbf{X}_t$  et de  $\dot{\mathbf{X}}_t$ :

$$\begin{cases} \mathbf{X}_{t+\delta t} = \mathbf{X}_t + \delta t \dot{\mathbf{X}}_t + \frac{\delta t^2}{2} \ddot{\mathbf{X}}_t, \\ \dot{\mathbf{X}}_{t+\delta t} = \dot{\mathbf{X}}_t + \delta t \ddot{\mathbf{X}}_t \end{cases} \quad (31)$$

Par la suite, on notera  $\mathbf{M}(k_t, \theta_t)$  par  $\mathbf{M}_t$ . Le modèle déformable global et ses dérivées temporelles développées au 1<sup>er</sup> ordre et au 2<sup>ème</sup> ordre s'expriment par:

$$\begin{cases} \mathbf{X}_t = \mathbf{M}_t [\bar{\mathbf{x}} + \Phi \mathbf{b}_t] + \mathbf{T}_t, \\ \dot{\mathbf{X}}_t = \dot{\mathbf{M}}_t [\bar{\mathbf{x}} + \Phi \mathbf{b}_t] + \mathbf{M}_t \Phi \dot{\mathbf{b}}_t + \dot{\mathbf{T}}_t, \\ \ddot{\mathbf{X}}_t = \ddot{\mathbf{M}}_t [\bar{\mathbf{x}} + \Phi \mathbf{b}_t] + \mathbf{M}_t \Phi \ddot{\mathbf{b}}_t + 2 \dot{\mathbf{M}}_t \Phi \dot{\mathbf{b}}_t + \ddot{\mathbf{T}}_t. \end{cases} \quad (32)$$

Par identification, on dérive les équations d'évolution temporelle des hyperparamètres du modèle:

$$\begin{cases} \mathbf{M}_{t+\delta t} = \mathbf{M}_t + \delta t \dot{\mathbf{M}}_t + \frac{\delta t^2}{2} \ddot{\mathbf{M}}_t, \\ \mathbf{T}_{t+\delta t} = \mathbf{T}_t + \delta t \dot{\mathbf{T}}_t + \frac{\delta t^2}{2} \ddot{\mathbf{T}}_t, \\ \mathbf{b}_{t+\delta t} = \mathbf{b}_t + \delta t \dot{\mathbf{b}}_t + \frac{\delta t^2}{2} \ddot{\mathbf{b}}_t + \tilde{\Phi}_{t+\delta t}^{-1} \mathbf{M}_t \Phi \ddot{\mathbf{b}}_t, \end{cases} \quad (33)$$

avec:

$$\tilde{\Phi}_{t+\delta t}^{-1} = [(\mathbf{M}_{t+\delta t} \Phi)^T (\mathbf{M}_{t+\delta t} \Phi)]^{-1} (\mathbf{M}_{t+\delta t} \Phi)^T. \quad (34)$$

Des équations similaires décrivent également le comportement de  $\dot{\mathbf{M}}_{t+\delta t}$ ,  $\dot{\mathbf{T}}_{t+\delta t}$  et  $\dot{\mathbf{b}}_{t+\delta t}$  (les termes croisés  $\dot{\mathbf{M}}_t \Phi \dot{\mathbf{b}}_t$  sont par ailleurs supposés négligeables). A ces équations d'évolution, on associe un filtre Kalman portant sur  $\mathbf{M}$ ,  $\mathbf{T}$  et  $\mathbf{b}$ . Les notations et équations du filtre sont précisés dans ce qui suit.

**Équations du filtre de Kalman** L'évolution dynamique du système est décrit par :

$$\mathbf{s}_{t+\delta t} = \mathbf{A}_t \mathbf{s}_t + \mathbf{B}_t \mathbf{u}_t + \xi_t \quad (35)$$

où  $\mathbf{s}_t$  est le vecteur d'état,  $\mathbf{u}_t$  est une entrée déterministe,  $\mathbf{A}_t$  est la matrice de transition et  $\xi_t$  est un bruit blanc gaussien centré de matrice covariance  $\mathbf{Q}_t = \mathbb{E} [\xi_t \xi_t^T]$ . La mesure  $\mathbf{w}_t$  est une fonction linéaire du vecteur d'état :

$$\mathbf{w}_t = \mathbf{H}_t \mathbf{s}_t + \eta_t \quad (36)$$

où  $\eta_t$  correspond à un bruit blanc gaussien centré de matrice de covariance  $\mathbf{V}_t = \mathbb{E} [\eta_t \eta_t^T]$  (dans notre cas,  $\mathbf{H}_t = (\mathbf{I}, \mathbf{0})$  et  $\mathbf{V}_t = \sigma_\eta^2 \mathbf{I}$ ).

Si  $\hat{\mathbf{s}}_{t_2|t_1}$  représente l'estimée minimisant l'erreur quadratique moyenne de  $\mathbf{s}_{t_2}$  connaissant les mesures jusqu'à  $t_1$  ( $t_1 \leq t_2$ ) et  $\mathbf{P}_{t_2|t_1}$  la matrice de covariance de l'erreur associée, les équations de remise à jour du filtre son :

$$\hat{\mathbf{s}}_{t|t} = \hat{\mathbf{s}}_{t|t-\delta t} + \mathbf{K}_t [\mathbf{w}_t - \mathbf{H} \hat{\mathbf{s}}_{t|t-\delta t}], \quad (37)$$

$$\mathbf{P}_{t|t} = [\mathbf{I} - \mathbf{K}_t \mathbf{H}] \mathbf{P}_{t|t-\delta t} \quad (38)$$

où le gain de Kalman est défini comme :

$$\mathbf{K}_t = \mathbf{P}_{t|t-\delta t} \mathbf{H}^T [\mathbf{H} \mathbf{P}_{t|t-\delta t} \mathbf{H}^T + \mathbf{V}_t]^{-1}, \quad (39)$$

Les équations de prédiction deviennent alors :

$$\hat{\mathbf{s}}_{t+\delta t|t} = \mathbf{A}_t \hat{\mathbf{s}}_{t|t} + \mathbf{B}_t \mathbf{u}_t, \quad (40)$$

$$\mathbf{P}_{t+\delta t|t} = \mathbf{A}_t \mathbf{P}_{t|t} \mathbf{A}_t^T + \mathbf{Q}_t \quad (41)$$

Nous avons développé deux filtres : un filtre à vitesse constante et un filtre reposant sur une estimée en ligne de la vitesse d'évolution des paramètres (filtre "à vitesse instantanée").

## 5.2 Filtre à vitesse constante

Soit  $\mathbf{z}_t$  représentant  $\mathbf{T}_t$ ,  $\mathbf{b}_t$  où  $\mathbf{M}_t$ . On définit le vecteur d'état  $\mathbf{s}_t = (\mathbf{z}_t \ \dot{\mathbf{z}}_t)^T$  et  $\mathbf{u}_t = 0$ . Les matrices de transition respectives pour les paramètres  $\mathbf{M}$  and  $\mathbf{T}$  d'une part, et  $\mathbf{b}$  d'autre part sont données par :

$$\mathbf{A}_t = \begin{pmatrix} \mathbf{I} & \delta t \mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}, \quad (42)$$

$$\mathbf{A}_t = \begin{pmatrix} \mathbf{I} & \delta t \mathbf{I} \\ \mathbf{0} & \mathbf{I} + \frac{3\delta t}{2} \tilde{\Phi}_{t+\delta t}^{-1} (\dot{\mathbf{M}}_t - \dot{\mathbf{M}}_{t+\delta t}) \Phi \end{pmatrix} \quad (42)$$

Les vecteurs de bruit d'état sont définis respectivement pour  $\mathbf{M}$  et  $\mathbf{T}$  d'une part, et  $\mathbf{b}$  d'autre part :

$$\xi_t = \begin{pmatrix} \frac{\delta t^2}{2} \ddot{\mathbf{z}}_t, \delta t \dot{\mathbf{z}}_t \end{pmatrix}^T, \quad (43)$$

$$\xi_t = \begin{pmatrix} \frac{\delta t^2}{2} \tilde{\Phi}_{t+\delta t}^{-1} \mathbf{M}_t \Phi \ddot{\mathbf{z}}_t \\ \left[ \delta t \mathbf{I} - \frac{\delta t^2}{2} \tilde{\Phi}_{t+\delta t}^{-1} \dot{\mathbf{M}}_{t+\delta t} \Phi \right] \tilde{\Phi}_{t+\delta t}^{-1} \mathbf{M}_t \Phi \dot{\mathbf{z}}_t \end{pmatrix} \quad (43)$$

On remarque que la prédiction et la remise à jour de  $\mathbf{b}$  dépend de  $\mathbf{M}$  (couplage entre paramètres). Nous avons adopté une formulation linéaire simple du filtre en découplant les filtres sur les paramètres  $\mathbf{M}$ ,  $\mathbf{T}$  et  $\mathbf{b}$  et en approchant  $\mathbf{M}_t$  par  $\mathbf{M}_{t|t}$  et  $\dot{\mathbf{M}}_{t+\delta t}$  par  $\dot{\mathbf{M}}_{t+\delta t|t}$ . Cette hypothèse donne des résultats satisfaisants en pratique. L'enchaînement des étapes de la procédure de suivi est le suivant : l'étape de prédiction permet de positionner le modèle à l'instant  $t+1$ , proche de la solution désirée. Les mesures sont obtenues par estimation, dans l'image à  $t+1$  au sens du MVM des hyperparamètres  $\Theta$ . Enfin, l'état est remis à jour selon l'équation 37. L'état ainsi filtré est le résultat final de l'estimation des déformations à l'instant  $t+1$ .

Ce filtre (de même que le filtre décrit dans la section suivante) est complété par une procédure de détection de rupture, permettant une réinitialisation du suivi lorsque le mouvement observé ne répond plus au modèle cinématique sous-jacent. Ce test est décrit de façon détaillée dans [11]. En pratique le nombre de ruptures reste faible (fig. 5).

## 5.3 Filtre "à vitesse instantanée"

Le vecteur d'état est défini ici par  $\mathbf{s}_t = \mathbf{z}_t$  et on introduit une entrée déterministe (vitesse instantanée)  $\mathbf{u}_t = \dot{\mathbf{z}}_t$  calculée par *différences finies* au cours du temps. Ce choix conduit à des estimations plus bruitées mais s'avère plus robuste lorsque la structure présente un comportement cinématique complexe (évolution du modèle à vitesse non constante).

Les matrices  $\mathbf{A}_t$  et  $\mathbf{B}_t$  sont identiques pour tous les paramètres :

$$\mathbf{A}_t = \mathbf{I} \quad \text{et} \quad \mathbf{B}_t = \delta t \mathbf{I}. \quad (44)$$

Les vecteurs de bruit d'état correspondent respectivement pour  $\mathbf{M}$  et  $\mathbf{T}$  d'un part, et  $\mathbf{b}$  d'autre part à :

$$\xi_t = \frac{\delta t^2}{2} \ddot{\mathbf{z}}_t, \quad \text{et} \quad \xi_t = \frac{\delta t^2}{2} \tilde{\Phi}_{t+\delta t}^{-1} \mathbf{M}_t \Phi \ddot{\mathbf{z}}_t. \quad (45)$$

$\xi_t$  intègre dans sa modélisation toutes les dérivées d'ordre supérieur assimilées à un bruit blanc. La procédure de suivi mise en œuvre est identique à celle présenté dans la section 5.2.

## 5.4 Résultats expérimentaux

L'approche proposée est illustrée ici sur des applications variées correspondant à des classes d'images et à des modèles de déformation différents.

Une première étude de cas concerne l'extraction et le suivi d'une main en mouvement dans un environnement texturé comportant des occlusions partielles (fig. 6). La séquence se compose de plus d'une centaine de trames. La figure 6a présente l'estimée finale des hyperparamètres au sens du MVM, à l'instant  $t-1$ . L'optimisation du processus de déformation à l'instant suivant  $t$  par des algorithmes déterministes rapides (ICM) ne garantit

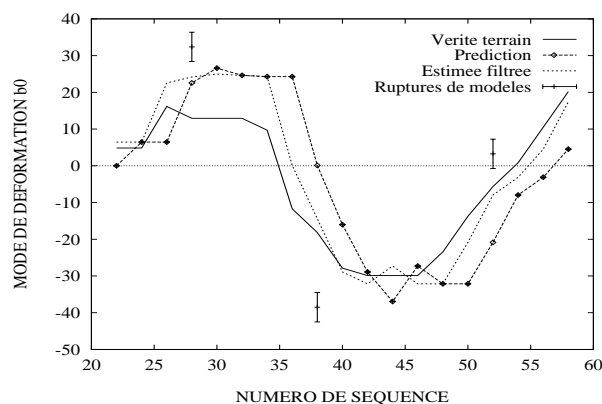


FIG. 5 - Suivi du 1<sup>er</sup> mode de déformation  $b_0$ .

pas dans ce cas un résultat satisfaisant (fig. 6b). Une optimisation des paramètres du modèle par des techniques stochastiques conduit à une segmentation satisfaisante (mais coûteuse!) de la main (voir fig. 6c – 4 mn de temps CPU sur une station Sun Sparc10).

La procédure de suivi permet d’obtenir une meilleure initialisation et donc de réduire notablement le temps de calcul par utilisation d’un algorithme d’optimisation déterministe (moins d’une minute de temps CPU) tout en maintenant une qualité de segmentation analogue. Les résultats présentés sur la figure 6 ont été obtenus par un suivi (avec un modèle à vitesse constante) des paramètres de déformation (des résultats similaires sont observés avec le filtre à vitesse instantanée). La figure 6d présente la configuration correspondant à la remise à jour de l’état du filtre (prise en compte de la mesure obtenue à l’aide d’algorithmes ICM d’optimisation des paramètres de déformation). On constate qu’on obtient une bonne segmentation finale, sans recours à une optimisation globale pour l’estimation des déformations.

La figure 5 présente l’évolution de la prédiction et de la remise à jour du 1<sup>er</sup> mode de déformation  $b_0$  sur une séquence synthétique d’une main en mouvement pour laquelle on connaît la “vérité terrain”. Les détections de rupture du modèle sont signalées par des barres verticales: le filtre est alors réinitialisé et une optimisation globale (recuit simulé) est utilisée sur l’image correspondante. Cette technique assure un suivi robuste en toutes circonstances dans la séquence.

Une seconde étude de cas concerne le suivi de lèvres en mouvement (fig. 7). Les gradients spatiaux représentent ici la seule information exploitée pour réaliser la segmentation. Dans ce cas, le filtre à vitesse instantanée s’est révélé plus robuste, car mieux adapté pour analyser le comportement cinématique complexe de la structure analysée. Les résultats de segmentation et suivi sont satisfaisants eu égard au caractère bruité des observations utilisées. La fig. 7c-d présente une autre application de ce modèle sur une séquence classique de télévision (“Claire”, séquence vidéophone), avec des performances similaires.

Une troisième étude de cas, présentée sur la figure 8, illustre la diversité du champ d’applications possible de cette technique de segmentation: il s’agit ici de suivre

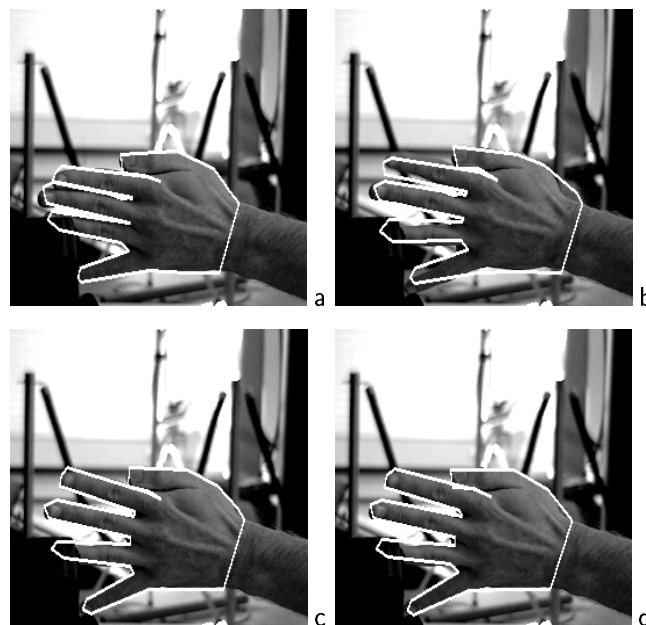


FIG. 6 - Extraction d’une main en mouvement; a) segmentation à  $t - 1$ ; b) segmentation par ICM à  $t$ ; c) segmentation par *recuit simulé* à  $t$ ; d) segmentation par ICM couplé au filtre de Kalman à  $t$ .

la boucle de déroulement d’un film cinématographique afin d’assurer un asservissement du moteur entraînant la bobine [27]. Des résultats très encourageants ont été obtenus (fig. 8) avec des temps de calcul relativement courts de l’ordre de 5 secondes par image. Le modèle de boucle dont le suivi est assuré ici apparaît en surimpression (noire), près du centre de la bobine.

## 6 Conclusion

Dans cet article, nous avons présenté un cadre statistique pour l’apprentissage et le suivi de structures non rigides dans des séquences d’images. L’approche présentée s’appuie une description statistique hiérarchique des déformations, associée à des techniques bayésiennes de segmentation et à un suivi par filtrage de Kalman des paramètres de déformation. Nous avons proposé une méthode non-supervisée pour réaliser l’apprentissage de la structure et des modes de déformation des formes analysées. Les performances atteintes sur des applications variées laissent entrevoir des perspectives prometteuses dans le domaine de la segmentation, du suivi et de l’interprétation de phénomènes déformables complexes.

## Références

- [1] A. AZARBAYEJANI, B. HOROWITZ, and A. PENTLAND. – Recursive estimation of structure and motion using relative orientation constraints. – In *Conf. Computer Vision and Pattern Recognition*, pages 294–299, New-York, June 1993.
- [2] B. BASCLE, P. BOUTHEMY, N. DERICHE, and F. MEYER. – Tracking complex primitives in an image sequence. – In *Int. Conf. on Pattern Recognition*, pages 426–431, Jerusalem, Israel, Oct. 1994.

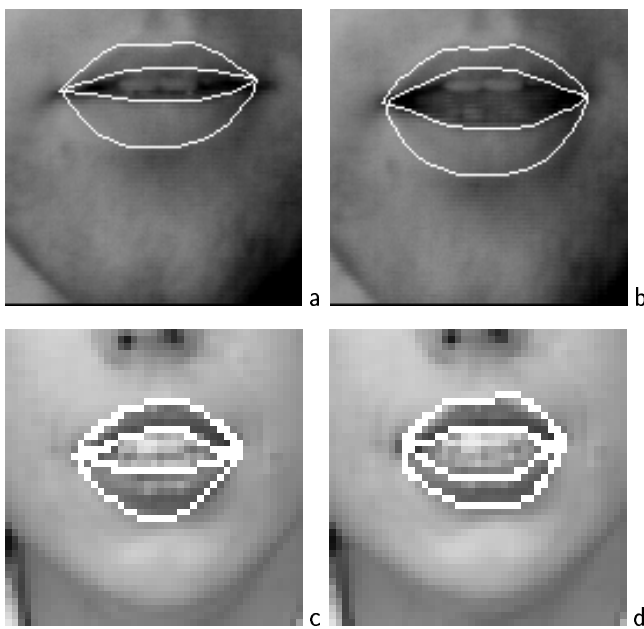


FIG. 7 - Suivi du mouvement des lèvres.

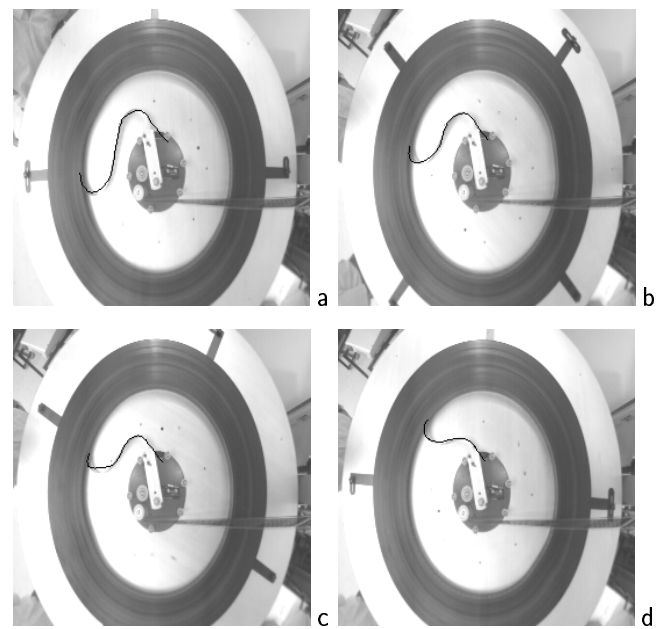


FIG. 8 - Suivi de la forme d'une boucle de déroulement (Laboratoire d'Automatique des Arts et Métiers).

- [3] M.O. BERGER. – *Les contours actifs : modélisation, comportement et convergence*. – PhD thesis, Institut National Polytechnique de Lorraine, Février 1991.
- [4] A. BLAKE, R. CURWEN, and A. ZISSERMAN. – A framework for spatiotemporal control in the tracking of visual contours. – *Int. J. Computer Vision*, Vol. 11, No 2: pp. 127–145, oct. 1993.
- [5] P. BOUTHEMY and A. BENVENISTE. – Modeling of atmospheric disturbances in meteorological pictures. – *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 6, No 5: pp. 587–600, Sept. 1984.
- [6] T.F. COOTES, C.J. TAYLOR, D.H. COOPER, and J. GRAHAM. – Training models of shape from sets of examples. – In *British Machine Vision Conf.*, pages 9–18, Leeds, UK, Sept. 1992.
- [7] T.F. COOTES, C.J. TAYLOR, D.H. COOPER, and J. GRAHAM. – Active shape models - their training and application. – *CVGIP: Image Understanding*, Vol. 61, No 1: pp. 38–59, 1994.
- [8] R. DERICHE and O. FAUGERAS. – Tracking line segments. – In *Proc. First European Conference on Computer Vision*, pages 259–268, Antibes, France, April 1990. Springer.
- [9] S. GEMAN and D. GEMAN. – Stochastic relaxation, Gibbs distributions and the bayesian restoration of images. – *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 6, No 6: pages 721–741, November 1984.
- [10] U. GRENANDER, Y. CHOW, and D.M. KEENAN. – *Hands. A Pattern Theoretic Study of Biological Shapes*. – Springer, 1991.
- [11] D. V. HINKLEY. – Inference about the change - point from cumulative sum - tests. – *Biometrika*, Vol. 58, No 3: pp. 509–523, 1971.
- [12] M. KASS, A. WITKIN, and D. TERZOPOULOS. – Snakes: Active contour models. – In *Proc. First Int. Conf. Comp. Vis.*, pages 259–268, London, UK, June 1987.
- [13] C. KERVRANN and F. HEITZ. – A hierarchical statistical framework for the segmentation of deformable objects in image sequences. – In *IEEE Conf. Comp. Vis. Patt. Rec.*, pages 724–728, Seattle, June 1994.
- [14] C. KERVRANN and F. HEITZ. – Robust tracking of stochastic deformable models in image sequences. – In *IEEE First Int. Conf. Image Processing*, Austin, USA, June 1994.
- [15] C. KERVRANN and F. HEITZ. – Learning structure and deformation modes of nonrigid objects in long image sequences. – In *International Workshop on Automatic Face- and Gesture Recognition*, Zurich, Switzerland, June 1995.
- [16] K.L. LAI and R.T. CHIN. – Deformable contours: modeling and extraction. – In *Conf. Computer Vision and Pattern Recognition*, pages 601–608, Seattle, USA, June 1994.
- [17] F. MEYER and P. BOUTHEMY. – Region-based tracking using affine motion models in long image sequences. – *CVGIP: Image Understanding*, Vol. 60, No 2:119–140, Sept. 1994.
- [18] A. MOHAMMAD-DJAFARI. – Estimation des hyperparamètres dans une approche bayésienne de la résolution des problèmes inverses linéaires. – In *Proc. Colloque GRETSI*, pages 201–204, Juan-Les-Pins, France, sept. 1993.
- [19] C. NASTAR and N. AYACHE. – Fast segmentation, tracking and analysis of deformable objects. – In *Proc. 4th Int. Conf. Comp. Vis.*, pages 275–279, Berlin, Germany, May 1993.
- [20] A. PENTLAND and B. HOROWITZ. – A modal framework for correspondence and recognition. – In *In Proc. 4th Int. Conf. Comp. Vis.*, pages 308–313, Berlin, Germany, May 1993.
- [21] A.P. PENTLAND and B. HOROWITZ. – Recovery of non-rigid motion and structure. – *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 13, No 7: pp. 730–742, July 1991.
- [22] N. ROUGON. – *Eléments pour la reconnaissance de formes tridimensionnelles déformables: application à l'imagerie*. – PhD thesis, Ecole Nationale Supérieure des Télécommunications, Février 1993.
- [23] L. H. STAIB and J. S. DUNCAN. – Boundary finding with parametrically deformable models. – *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 14, No 11: pp. 1061–1075, Nov. 1992.
- [24] D. TERZOPOULOS and D. METAXAS. – Dynamic 3D models with local and global deformations: Deformable superquadrics. – *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 13, No 7: pp. 703–714, July 1991.
- [25] D. TERZOPOULOS and A. WITKIN. – Deformable models. physically based models with rigid and deformable components. – *IEEE Computer Graphics and Applications*, pages 41–51, Nov. 1988.
- [26] A.L. YUILLE. – Feature extraction from faces using deformable templates. – *Int. J. Computer Vision*, Vol. 8, No 2: pp. 99–111, 1992.
- [27] D. ZUGAJ and V. LATTUATI. – Contrôle en temps réel de la forme d'une boucle de déroulement par vision artificielle. – In *QCAV' 95*, Le Creusot, France, 1995.