

**Segmentation spatio-temporelle  
et suivi dans une séquence d'images**

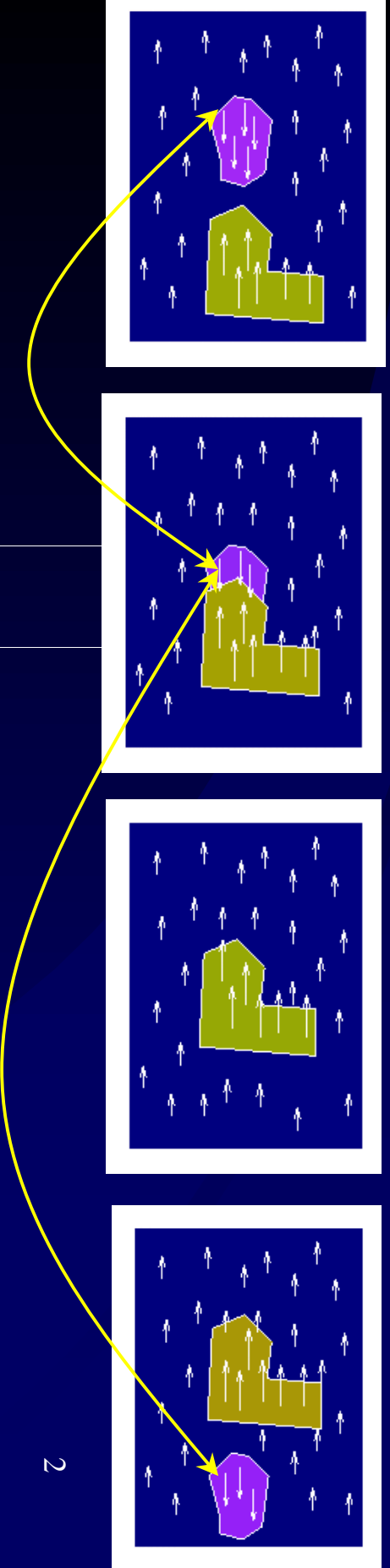
**Application à la structuration et à l'indexation de vidéo**

**Marc Gelgon**

**Projets Temis/Vista - IRISA**

# Segmentation, suivi, trajectographie

- Partitionnement d'image en régions pertinentes
  - segmentation au sens du mouvement
- Extension temporelle de ce partitionnement
  - suivi court-terme de la partition
  - trajectographie des régions



# Plan de l'exposé

## Vision dynamique (partie 1)

- 1) Segmentation au sens du mouvement, suivi
- 2) Trajectographie de régions

# Application : l'indexation de vidéo

- Consultation plus efficace de documents vidéo
- Nouvelles formes d'accès au contenu
- Nécessité d'une analyse préalable  
(structuration en éléments pertinents)

# Plan de l'exposé

## Vision dynamique (partie 1)

- 1) Segmentation au sens du mouvement, suivi
- 2) Trajectographie de régions

## Structuration et indexation (partie 2)

- 1) Découpage de vidéo en plans  
Structuration spatio-temporelle des plans
- 2) Résumés vidéo et base d'objets mobiles

## 1.1

# Segmentation au sens du mouvement et suivi de partition

# Segmentation au sens du mouvement

- Un problème fondamental
  - analyse, interprétation
  - codage vidéo
- Un problème difficile
  - interdépendance mesure du mouvement / segmentation
  - écart entre critère de segmentation et interprétation sémantique

# Segmentation au sens du mouvement : quelques grands axes

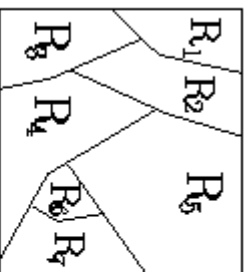
- Extraction séquentielle des régions [Irani92]
- Conjointe
  - loi de mélanges [Ayer96, Weiss96]
  - étiquetage statistique contextuel [Odobez94, Stiller97]
- Regroupement de régions élémentaires
  - régions intensité/couleur, blocs [Garcia-Garduno94, Wang94]



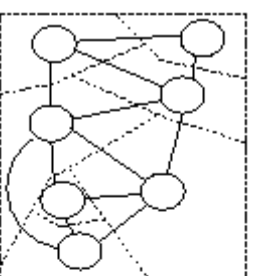
# Segmentation au sens du mouvement : approche proposée (1)

- Partition spatiale et partition mouvement
- Partition spatiale :
  - sortie utile de l'algorithme
  - moyen d'obtenir une partition au sens du mouvement
- Etiquetage statistique :
  - pixel (partition spatiale) et graphe de régions (mouvement)

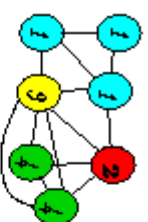
# Segmentation au sens du mouvement : approche proposée (2)



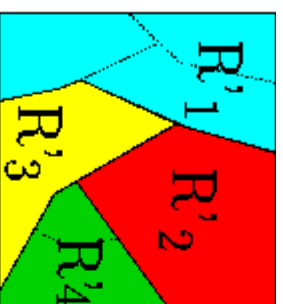
**Partition spatiale**  
(critère d'intensité,  
couleur ou texture )



**Graphes d'adjacence  
des régions élémentaires**



**Étiquetage au sens du mouvement  
des noeuds du graphe**



**Partition au sens du mouvement**

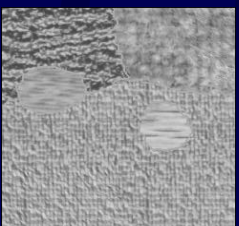
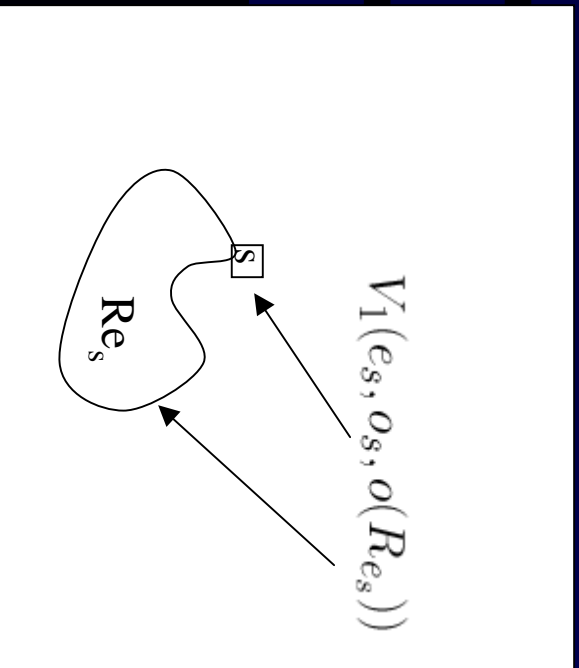
# Segmentation spatiale

- Problème d'étiquetage statistique bayésien :

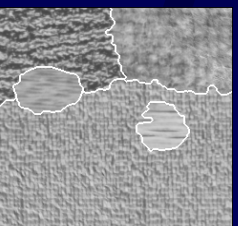
$$\hat{e} = \arg \min_{e \in \Omega} U_1(e, o) + U_2(e)$$

- Fonction d'énergie :

$$U(e, o) = \sum_{s \in S} V_1(e_s, o_s, o(R_{e_s})) + \sum_{\langle s, s^* \rangle \in C} V_2(s, s^*)$$



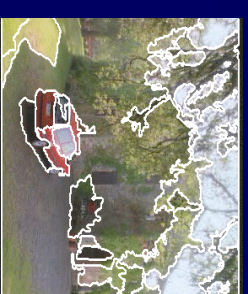
Texture



Intensité



Couleur



[Kervrann95]

- Détermination en ligne du nombre de régions

# Segmentation au sens du mouvement :

## Etiquetage statistique sur un graphe

Objectif : groupes de régions cohérents au sens du mouvement

- Champ d'étiquettes  $e'$  sur des nœuds du graphe des régions
- Approche bayésienne-markovienne
- Définition et minimisation d'une fonction d'énergie

$$\hat{e}' = \arg \min_{e' \in \Omega'} U'(e', o')$$

$$U'(e', o') = U_1'(e', o') + U_2'(e') + U_3'(e', o')$$

similarité de mouvement

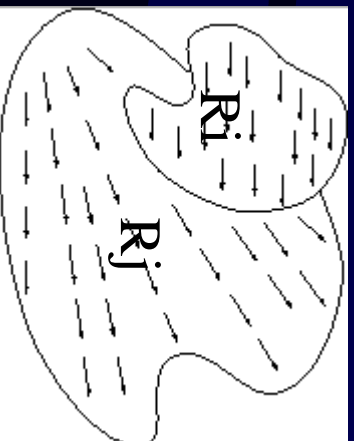
géométrie des régions

incertitude sur le mouvement

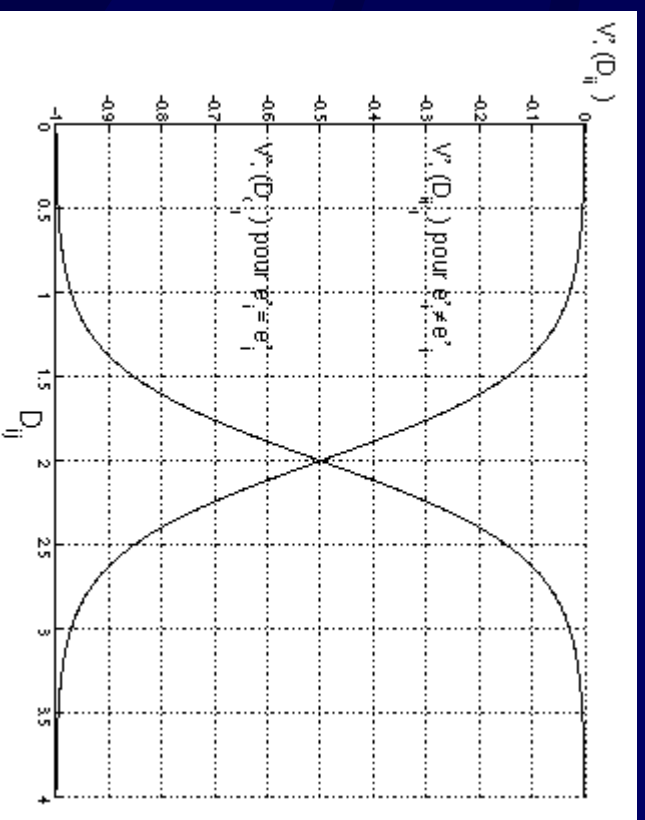
- $U_1', U_2', U_3'$  définis par des potentiels locaux  $V_1', V_2', V_3'$

# Segmentation au sens du mouvement : Définition de la fonction d'énergie $U'$

- Estimation robuste d'un modèle de mouvement par région spatiale
- Cohérence de mouvement  $D_{ij}$  définie pour une paire de régions voisines  $(R_i, R_j)$
- Potentiel  $V'_1$  défini sur des cliques binaires  $(s_i, s_j)$



Différence entre champs de vitesse  
à la frontière  $(R_i, R_j)$

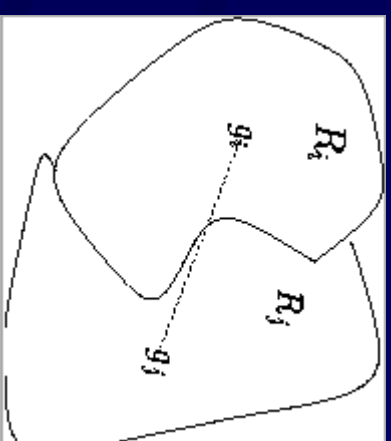


Potentiel  $V'_1(s_i, s_j) = \text{fct}(D_{ij})$

# Segmentation au sens du mouvement :

Définition et minimisation de la fonction d'énergie  $U'$

- $V'_2$  : favoriser des regroupements à géométrie compacte

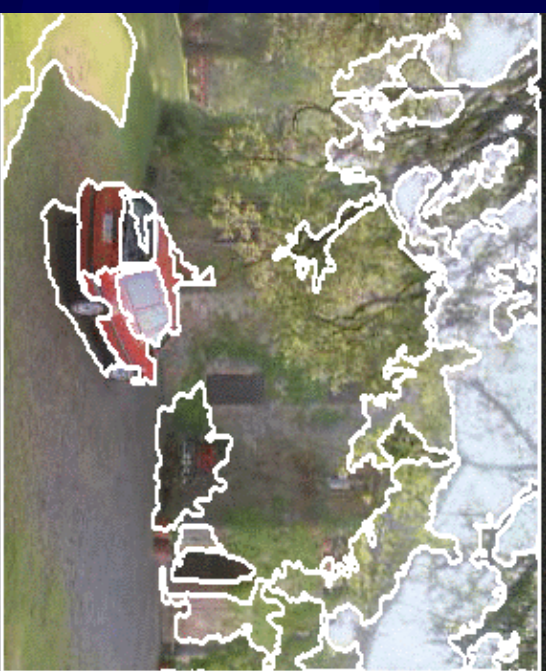


- $V'_3$  : prise en compte explicite des régions dont le mouvement est mal estimé.
- Minimisation HCF modifiée

# Résultats expérimentaux : séquence ajax



Image originale



Segmentation spatiale



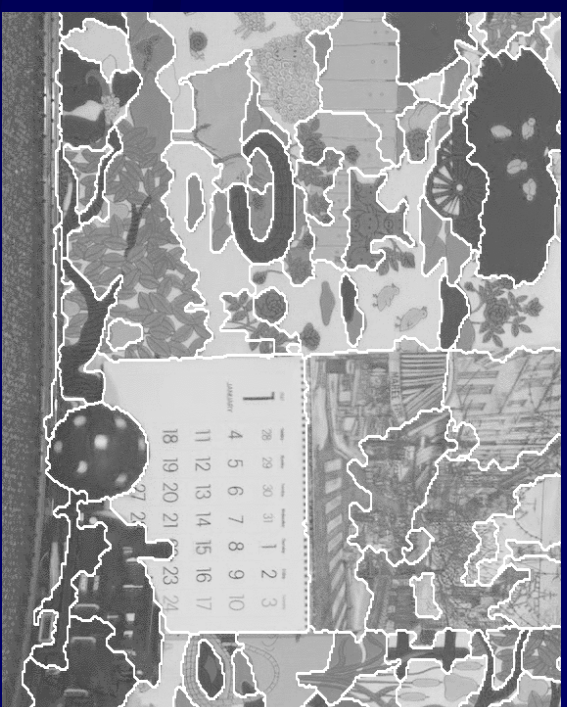
Segmentation mouvement

Source : INA

# Résultats expérimentaux : séquence mobi



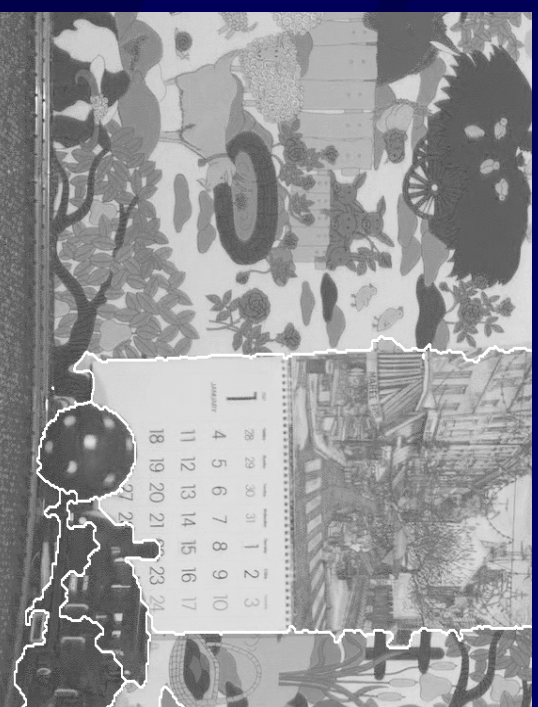
Image originale



Segmentation spatiale



Segm. Mouvement (sans U<sup>3</sup>)

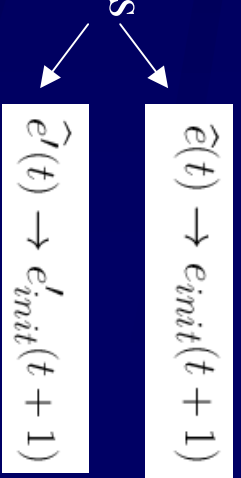


Segm. Mouvement (avec U<sup>3</sup>)



# Suivi temporel des deux partitions

- Principe de propagation temporelle / mise à jour de  $e$  et  $e'$  [François90]
- Pour chaque problème d'étiquetage
  - construction d'une configuration initiale d'étiquettes
  - mise à jour des frontières et du nombre de régions
- configuration initiale proche d'un minimum local (pour  $U$  et  $U'$ )
  - convergence rapide
  - conservation de l'étiquette, lien temporel établi



# Suivi temporel : résultats

(séquence interview)

# Caractéristiques de l'approche proposée

- Qualités :
  - possible remise en question des regroupements  
(≠ clustering, fusion)
  - les deux partitions sont disponibles à chaque instant
  - possibilité d'introduire d'information contextuelle
  - exploitation des frontières d'intensité
- Limites :
  - hypothèse sur la hiérarchie spatial/mouvement
  - fiabilité de l'estimation du mouvement sur les petites régions
  - « mémoire courte » de l'algorithme

1.2

# Trajectographie de régions

# Estimation et association de trajectoires exemple de problème:



Séquence originale

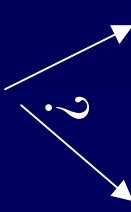


Cartes de segmentation

- Lien-temporel fourni par suivi court-terme, occultations partielles
- Association de trajectoires partielles [Meyer'93].

# Trajectographie de régions :

## formulation du problème

- M modèles de trajectoires  $X = \{ \dots, \{x_1(t), \dots, x_M(t)\}, \dots \}$
  - à estimer à partir des mesures  $Z = \{ \dots, \{z_1(t), \dots, z_{st}(t)\}, \dots \}$
- 

- M trajectoires partielles, nombre de trajectoires réelles ( $< M$ ) à déterminer

- Mesures disponibles :
  - masque de la région
  - étiquette (lien court-terme)
  - modèle de mouvement

$$x_m(t) = \left[ \mathcal{G}_m(t), \Theta_m(t) \right]^T$$

$$z_j(t) = \left[ \tilde{\mathcal{G}}_j(t), \tilde{\Theta}_j(t) \right]^T$$

# Trajectographie de régions :

Vecteurs d'états et de mesures : composante cinématique

Estimation des paramètres de mouvement

$$\Theta_m(t) = [a_{m1}^1(t), \dots, a_{m6}^6(t)]^T \quad \text{avec}$$

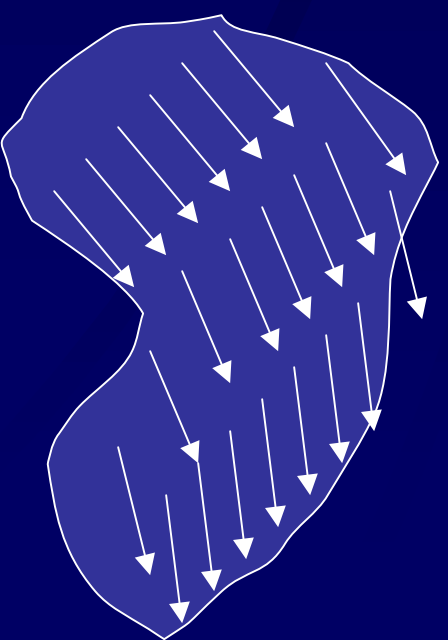
$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \begin{pmatrix} a_0 + a_2 \cdot x + a_3 \cdot y \\ a_1 + a_4 \cdot x + a_5 \cdot y \end{pmatrix}$$

Modèle d'évolution temporelle

$$\begin{bmatrix} a \\ \dot{a} \end{bmatrix}_m^r(t+1) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ \dot{a} \end{bmatrix}_m^r(t) + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}_m^r(t)$$

Modèle de mesure

$$\tilde{a}_m^r(t) = a_m^r(t) + \eta_m^r(t)$$



# Trajectographie de régions :

Vecteurs d'états et de mesures : composante géométrique

Modèle géométrique polygonal

$$G_m(t) = \{P_m^1(t), \dots, P_m^n(t)\}$$

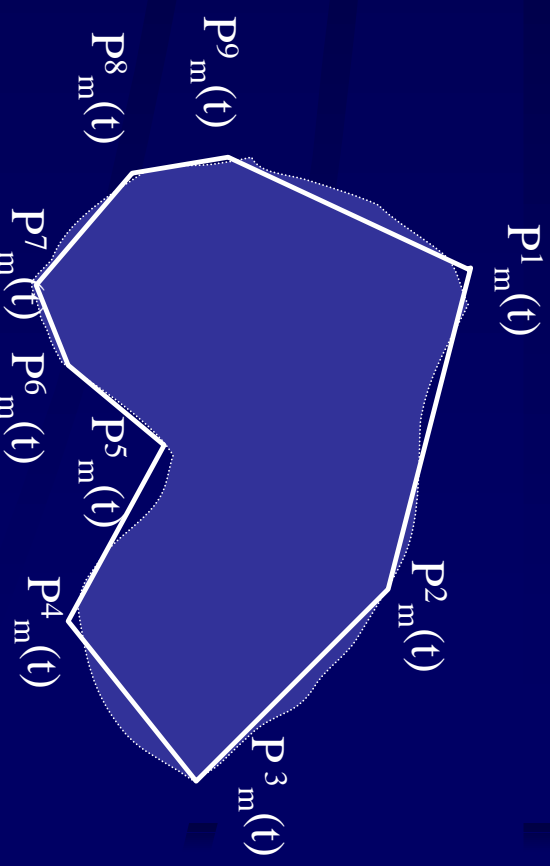
$$P_m^r(t) = \begin{bmatrix} u \\ v \end{bmatrix}_m^r(t)$$

Modèle d'évolution temporelle

$$\begin{bmatrix} u \\ v \end{bmatrix}_m^r(t+1) = \begin{bmatrix} 1+a_2 & a_3 \\ a_4 & 1+a_5 \end{bmatrix}_m(t) \begin{bmatrix} u \\ v \end{bmatrix}_m^r(t) + \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}_m(t) + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}_m^r(t)$$

Modèle de mesure

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \end{bmatrix}_m^r(t) = \begin{bmatrix} u \\ v \end{bmatrix}_m^r(t) + \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}_m^r(t)$$





# Trajectographie de régions :

## Choix de la méthode PMHT

- Inter-dépendance des problèmes d'estimation et d' association
- approches classiques : MHT [Cox96] (combinatoire), JPDAF
- PMHT : variables d'association : données manquantes ( un problème d'estimation globale [Streit93] )
- méthode «batch»

# Trajectographie de régions :

## Estimation et association conjointes

$$p(\mathcal{Z} | \mathcal{X}, \Pi) = p(X(0)) \prod_{t=1}^T \underbrace{p(Z(t) | X(t), \Pi(t))}_{\text{vraisemblance des mesures}} \underbrace{p(X(t) | X(t-1))}_{\text{evolution a priori}}$$

$$\prod_{t=1}^T p(Z(t) | X(t), \Pi(t)) = \prod_{t=1}^T \prod_{j=1}^{s_t} \prod_{m=1}^M p(z_j(t) | x_m(t)) \pi_m(t)$$

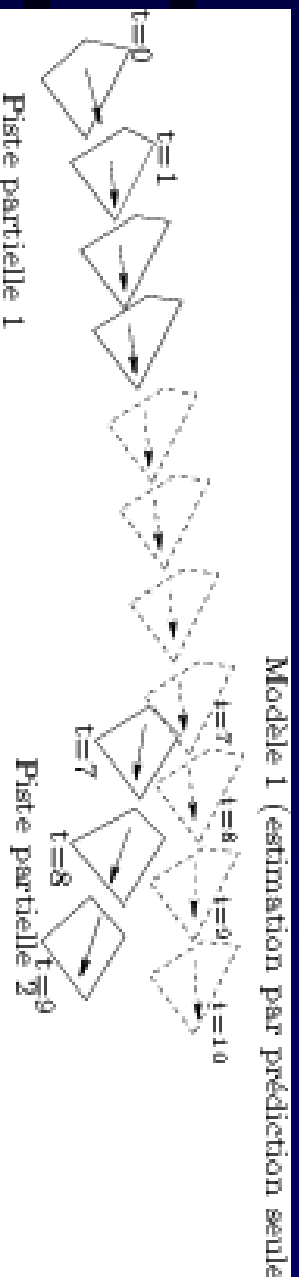
[Gauvrit-Le Cadre97]

- Estimation par filtrage de Kalman et algorithme EM
- Association pistes partielles-modèles [Giannopoulos97]

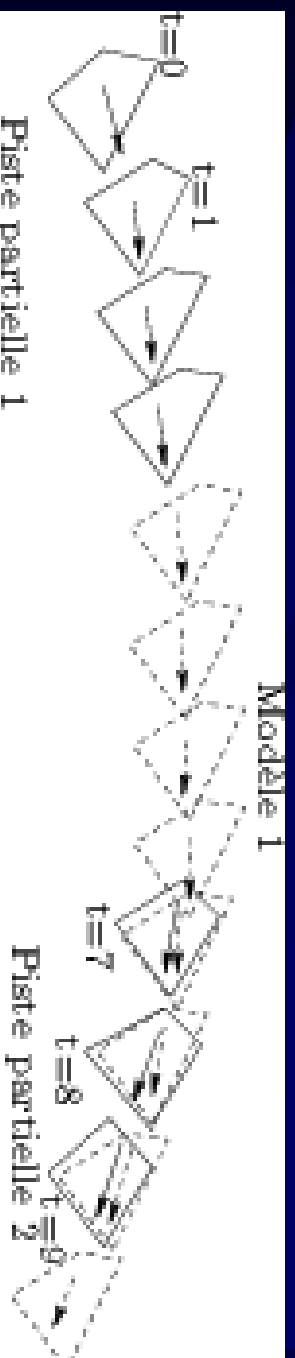
# Trajectographie de régions :

## Evolution des états et affectations

Etats initiaux

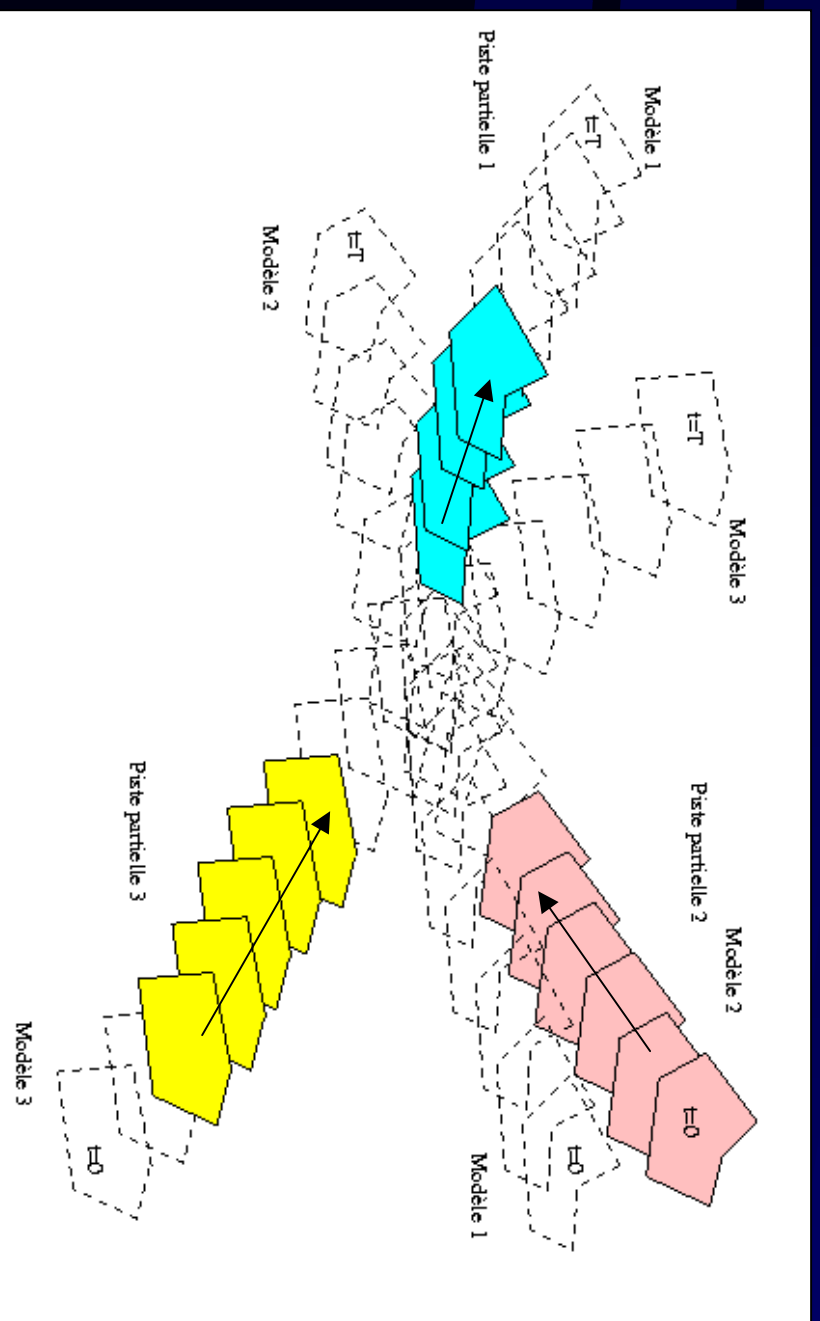


Si adéquation suffisante modèle 1 / piste partielle 2 au sens de  $p(z_j(t) | x_m(t))$



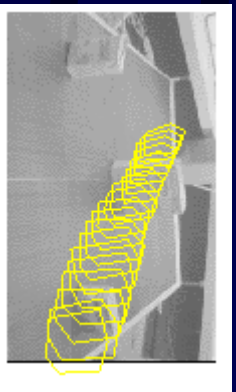
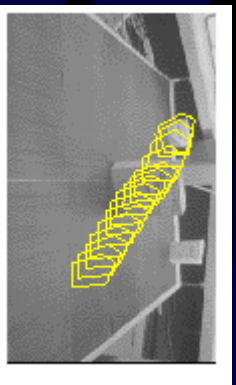
# Trajectographie de régions : Adaptation au suivi des régions

- Procédure efficace pour l'initialisation
- Etape de suppression des mesures occultées
- Estimation initiale de l'état à chaque instant

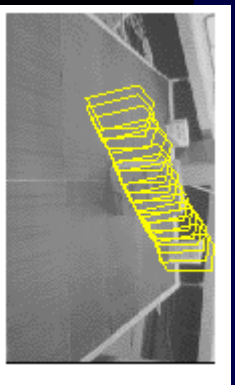


# Résultat expérimental

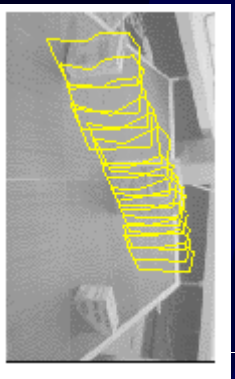
Etats (initialisation)



Modèle 1



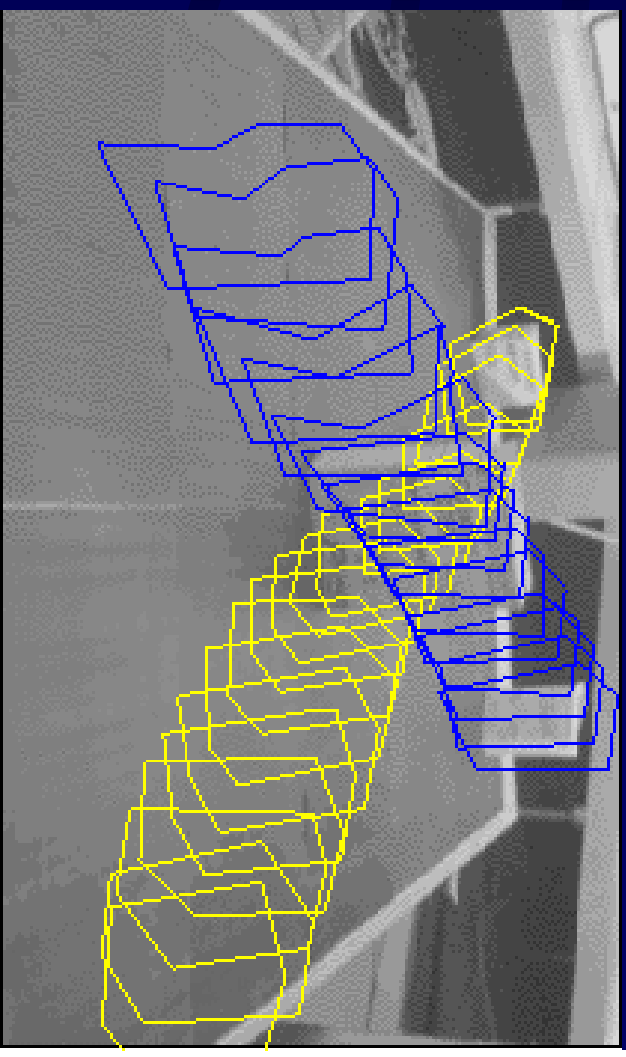
Modèle 2



Modèle 3

Modèle 4

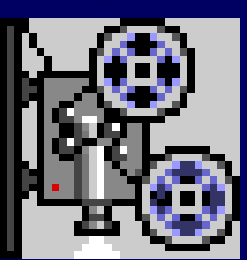
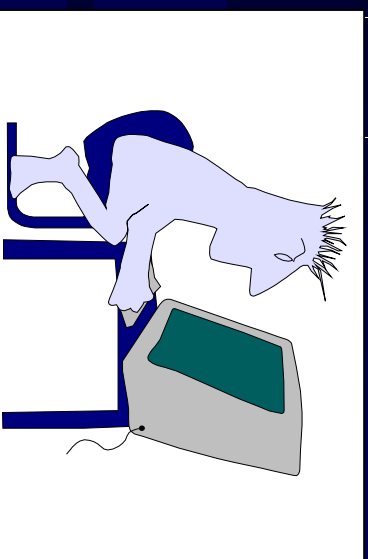
Etats (convergence)



# 2

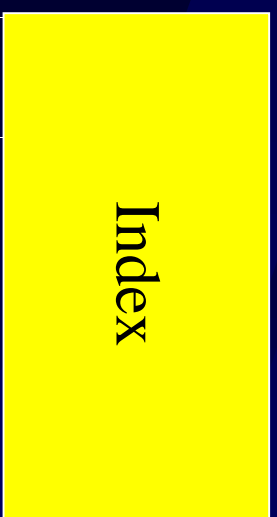
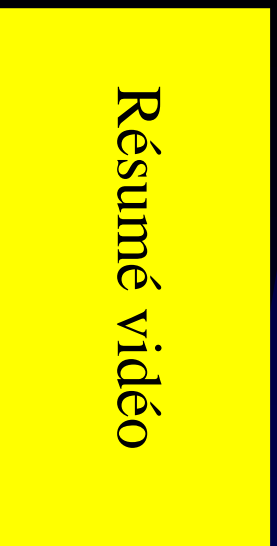
## Structuration de vidéo pour l'indexation

# Outils pour la consultation efficace des documents vidéo



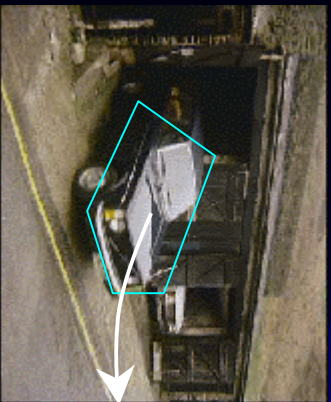
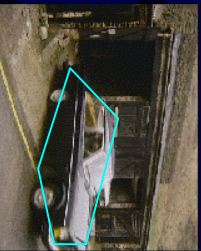
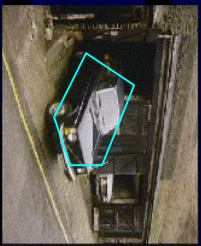
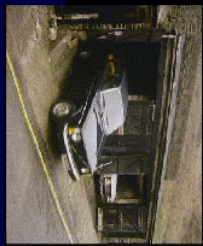
Recherche  
ordinaire

Requêtes



Analyse préalable du contenu

# Schéma synoptique du système proposé



Droite



Avant



Droite

Résumé du plan

Base d'éléments mobiles



## 2.1

# Découpage de vidéo en plans

# Découpage de vidéo en plans

- Approches classiques exploitent l'intensité
- Analyse directe du flux MPEG
- Difficultés :
  - transitions progressives (fondus, volets)
  - mouvement de caméra
  - présence d'éléments mobiles
  - effets d'illumination

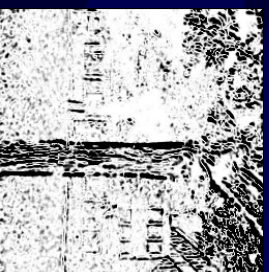
# Découpage en plans : approche proposée

- Exploitation d'une information « intrinsèque » (mouvement dominant apparent)
- Estimation robuste du modèle affine 2D de ce mouvement
- $\gamma(t)$  : taille (normalisée) du support associé [Bouthemy96]
- Evolution temporelle de  $\gamma(t)$
- Détection des changements de plans par test de Hinkley sur  $\gamma(t)$



Dans un plan

$$\gamma(t) \approx 1$$



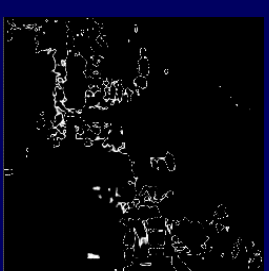
Elément non-conforme

$$\gamma(t) \approx 0.7$$



Fondu

$$\gamma(t) \approx 0.4$$



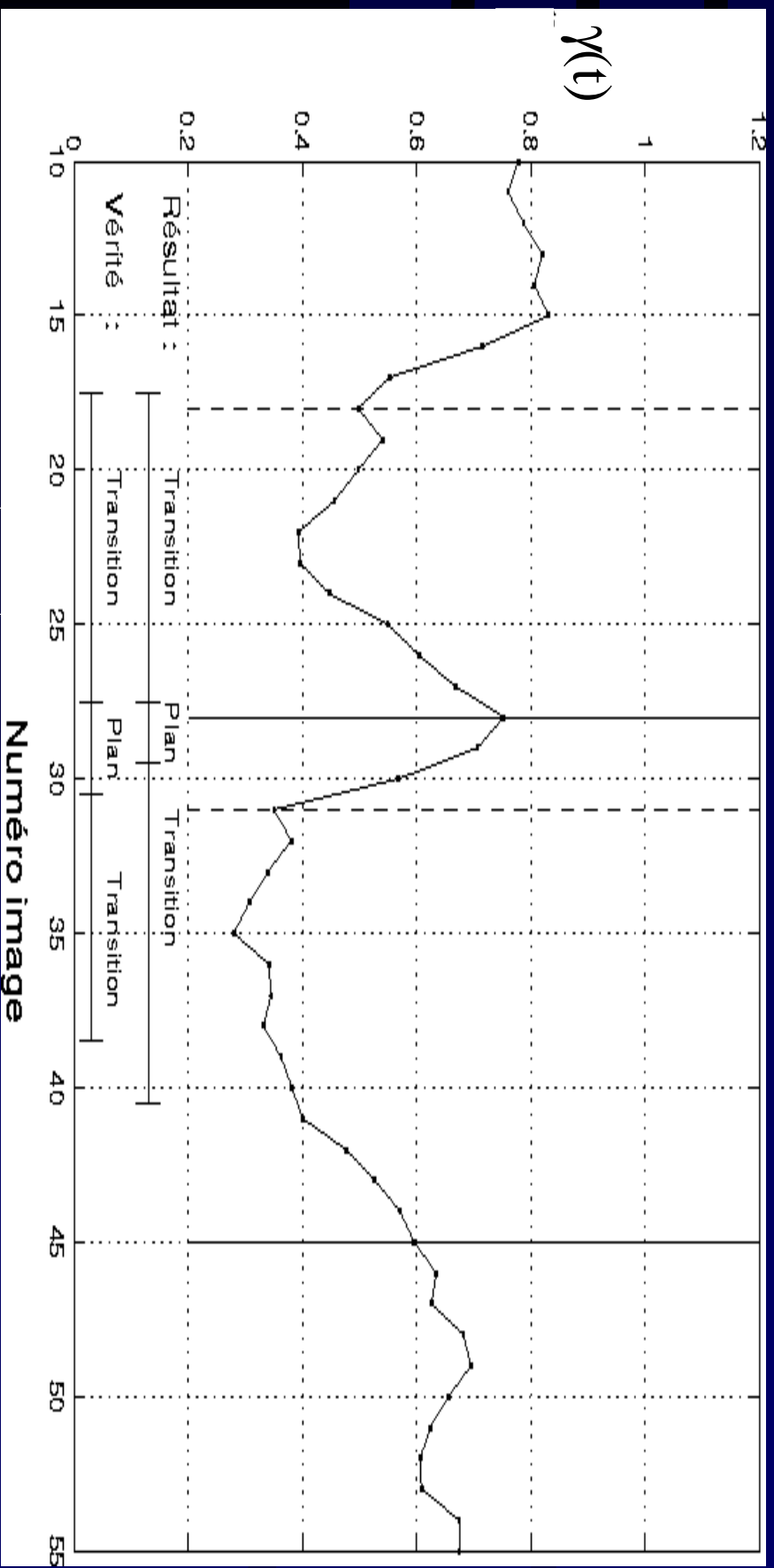
Cut

$$\gamma(t) \approx 0$$

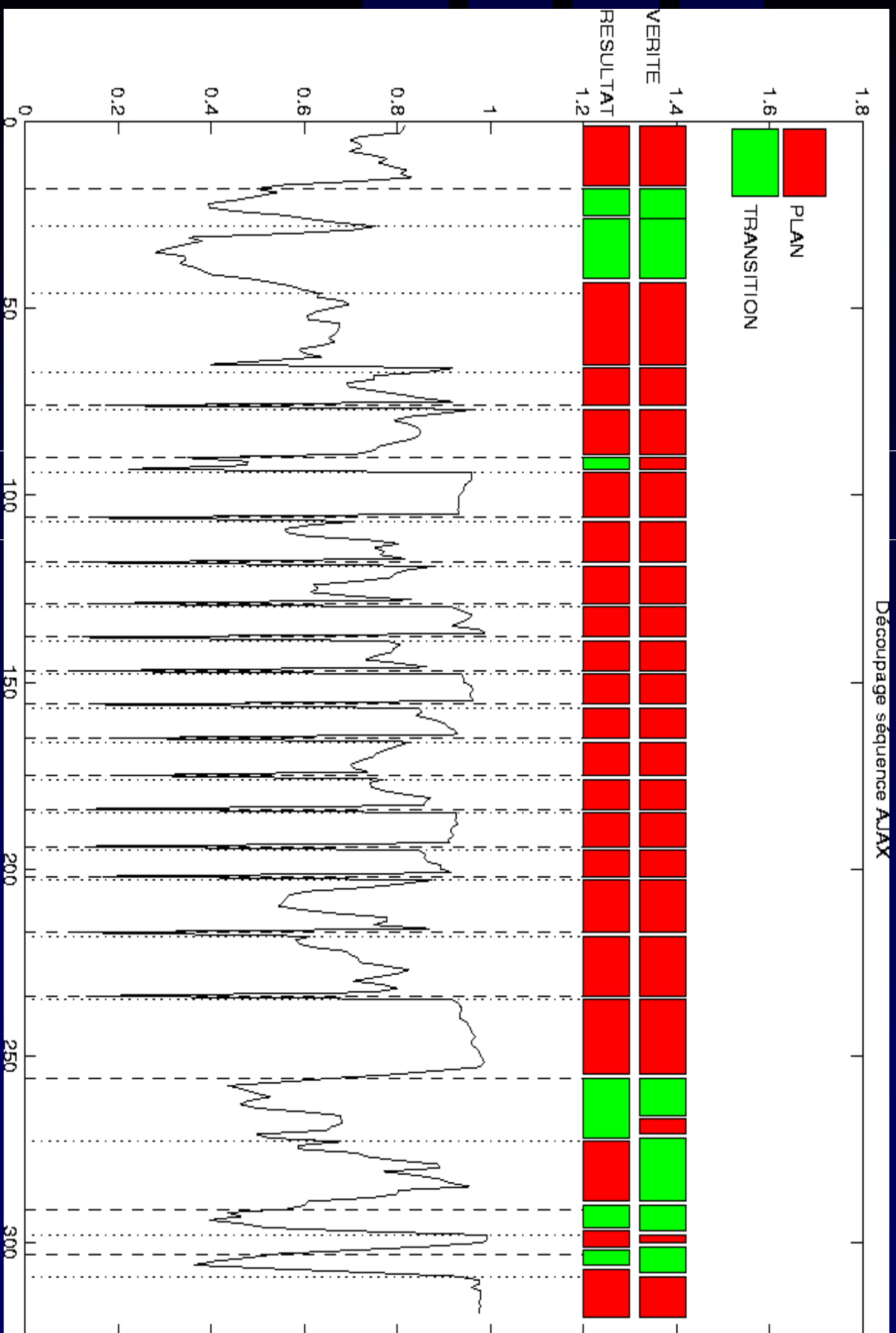
# Découpage en plans : approche proposée

Détection de changement de plan :

détection de saut sur  $\gamma(t)$  par test cumulatif de Hinkley



# Résultat sur la séquence Ajax



## 2.2

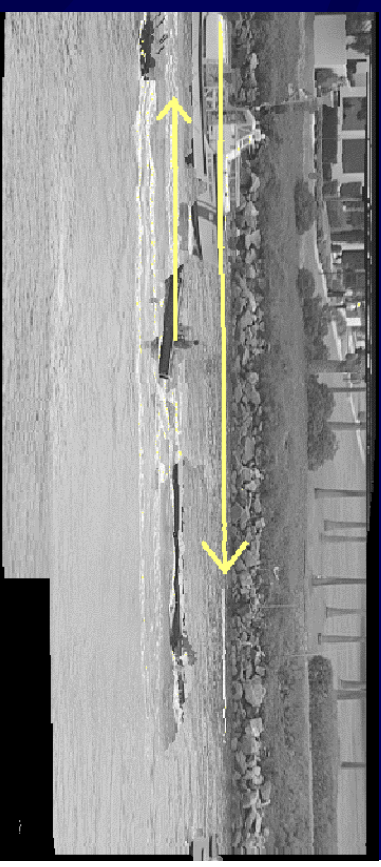
# Résumé vidéo et base d'objets mobiles

# Vue synoptique du contenu d'un plan

- Création d'une image mosaïque du plan
  - choix d'une image de référence
  - recalage de toutes les images dans le repère défini par cette image (compensation du mouvement global dominant)



- Contenu dynamique : trajectoire des éléments mobiles extraits
  - extraction / suivi d'éléments mobiles
  - trajectoire superposée sur la mosaïque



# Descripteurs de mouvement

- Qualification du mouvement de la caméra [François90]  
(plan fixe, panoramique, travelling)
- Descripteurs qualitatifs de mouvement des éléments extraits
  - sans compensation du mouvement de la caméra
  - avec compensation

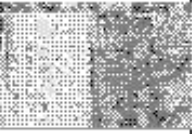


Indexation sur des informations du mouvement



# Résultat expérimental (séquence mobi)



Mosaïque et trajectoire des  
éléments mobiles extraits

Nom	Calendrier	Train	Ballon
Icône			
Image apparition	1	1	1
Image disparition	125	125	125
Mouvement apparent	haut-gauche (1-56) bas-gauche (60-125)	gauche (1-125)	gauche, rotation - (1-33) droite (34-46) gauche, rotation - (47-125)
Mouvement "vue globale"	haut (1-56) bas (60-125)	gauche (1-125)	gauche, rotation - (1-33) statique (34-46) gauche, rotation - (47-125)

Base d'éléments mobiles

# Extraction interactive de zone

- Limitations des schémas « tout automatique » pour l'indexation vidéo
- Approche semi-automatique :
  - détournage initial de la région
  - suivi par estimation de transformations affines
  - alarme en cas de décrochage du suivi
- Résultats expérimentaux

Manuel

Automatique

# Conclusion et perspectives

# Contributions

- Segmentation / suivi de partitions :
  - définition d'une technique « mixte » pixels / régions
  - étiquetage statistique de graphe
  - deux partitions (spatial, mouvement) à chaque instant avec lien temporel
- Trajectographie :
  - choix d'une méthode PMHT et adaptation aux primitives régions
  - technique d'identification des modèles redondants.
- Indexation vidéo :
  - construction d'un système complet de structuration spatio-temporelle
  - indexation sur des informations de mouvement
  - prise en compte d'un contexte d'utilisation semi-automatique

# Perspectives

- Segmentation / suivi :
  - transposition du paradigme proposé pour la détection du mouvement
  - prise en compte de la cohérence temporelle (batch)
- Trajectographie :
  - prise en compte de critères de type « image »
  - application à la correction des « sursegmentations »
- Indexation :
  - indexation par des descripteurs statiques (liens « hyper-vidéo »)
  - adaptation de la complexité de la structuration / du résumé au contenu