Approche spatio-temporelle pour l'analyse de séquences d'images. Application en détection de mouvement

Spatiotemporal Approach for image sequences analysis. Application to Motion Detection

par Alice CAPLIER, Franck LUTHON

Laboratoire de Traitement d'Images et de Reconnaissance de Formes
Institut National Polytechnique de Grenoble
LTIRF, INPG, 46 avenue Félix-Viallet
38031 Grenoble Cedex, France
Tel : 04 76 57 43 72 Fax : 04 76 57 47 90
Email : luthon@tirf.inpg.fr

résumé et mots clés

Dans cet article, on propose une nouvelle stratégie pour aborder le problème de l'analyse du mouvement dans les séquences d'images. L'originalité de l'approche consiste à considérer une séquence d'images non pas comme une succession d'images 2D mais comme un flux de données à trois dimensions (x,y,t).
En appliquant cette stratégie au problème de la détection de mouvement dans des séquences d'images acquises avec une caméra fixe, on définit un modèle markovien spatio-temporel associé à une relaxation spatio-temporelle. Grâce à une modélisation fine des interactions dans le voisinage 3D pris en compte, on obtient des résultats intéressants pour détecter des objets dans des séquences d'images très bruitées ou des objets dont le déplacement est faible d'une image à l'autre.
Pour améliorer les performances de l'algorithme lors de la détection d'objets peu texturés ou de mouvement sous-pixel, on peut également l'intégrer dans une structure multirésolution pour laquelle la hiérarchisation des données provient de filtrages passe-bas et de sous-échantillonnages dans chacune des trois dimensions x, y et t. Des résultats obtenus sur des séquences synthétiques et naturelles montrent l'intérêt de cette approche.

Détection de mouvement, séquence d'images, champ de Markov, approche spatio-temporelle, multirésolution.

abstract and key words

In this paper, a new strategy for motion analysis in image sequences is proposed.
The originality of this work is that an image sequence is seen as a 3D data flow instead of a series of 2D-images. By applying such an approach to the problem of motion detection in image sequences acquired with a static camera, a spatiotemporal Markovian model associated with spatiotemporal relaxation is defined. Thanks to an adequate modelling of spatiotemporal interactions beetween pixels belonging to a 3D neighbourhood, good results may be obtained for detecting moving objects in noisy sequences or slow moving objects.
In order to improve the ability of the algorithm to detect poorly textured objects or subpixel moving objects, it may be integrated in a multiresolution scheme. The data pyramid is built by using 3D low-pass filters and 3D subsamplings. Results on synthetic and real worl image sequences are exhibited, showing the validity of this approach.

Motion detection, image sequences, Markov Random Field (MRF), spatiotemporal approach, multiresolution.

1. introduction

Plusieurs travaux relatifs à l'analyse de séquences d'images ont mis en évidence l'intérêt de développer des algorithmes spatio-temporels [1]. Les méthodes proposées s'efforcent d'utiliser au mieux l'information temporelle.

Dans cet article, on présente une approche spatio-temporelle originale pour l'analyse de séquences d'images. Elle consiste à considérer une séquence d'images non pas comme une succession d'images 2D mais comme un flux de données à trois dimensions (x, y, t). L'intérêt de cette approche est illustré dans le cas de la détection de mouvement, c'est-à-dire de la localisation des zones fixes et mobiles dans les séquences d'images acquises avec une caméra fixe. La détection de mouvement est vue comme un problème d'étiquetage binaire. Pour résoudre cette tâche, on utilise le formalisme des champs aléatoires de Markov qui est particulièrement bien adapté à la résolution de problèmes de segmentation d'images statiques [5, 10] et dynamiques [6].

Les premiers travaux relatifs à la détection de mouvement selon une approche markovienne [6, 13] ont conduit au développement d'un modèle intégrant l'information de mouvement contenue dans 2 ou 3 images successives (instants passé, présent et futur). Néanmoins, l'expérience montre que la prise en compte de deux ou trois images seulement s'avère insuffisante pour détecter les objets mobiles dans certains cas tels que l'analyse de séquences très bruitées, la détection d'objets mobiles peu texturés ou d'objets de mouvement très lent. Pour pallier les déficiences des algorithmes liées à une analyse de la scène sur un horizon temporel restreint, la démarche classique est d'extraire des informations de mouvement de meilleure qualité comme c'est le cas par exemple dans [15]. La structure générale du modèle markovien associé à l'algorithme est classique. En revanche, la définition des observations associées à ce modèle a été particulièrement soignée afin d'améliorer les performances globales de l'algorithme de détection de mouvement.

Le travail présenté s'appuie sur le fait que l'analyse du mouvement est plus robuste si on tient compte des informations contenues dans un plus grand nombre d'images consécutives et que, lors de son déplacement, un objet mobile décrit un volume dans l'espace (x, y, t). Il est donc pertinent de s'intéresser à la trace de l'objet dans ce volume. Cependant le problème de l'occultation d'objets mobiles n'est pas abordé ici car l'algorithme proposé est un algorithme de détection de mouvement et non de segmentation au sens du mouvement. De ce fait, l'information extraite est la présence ou non de mouvement et non le suivi au cours du temps des divers objets d'une scène.

Après de brefs rappels sur la théorie des champs de Markov (section 2), l'approche spatio-temporelle proposée, appliquée au cas de la détection de mouvement, conduit à la définition d'un champ de Markov **spatio-temporel** associé à une relaxation sur un volume borné de données (section 3). La démarche consiste à améliorer le modèle a priori associé au champ d'étiquettes en gardant des observations relativement rudimentaires. Il suit une analyse critique des performances de cet algorithme tant du point de vue de la qualité des résultats que de la complexité calculatoire (section 4). Une hiérarchisation des données sous la forme d'une pyramide de séquence est proposée dans la section 5. En associant l'algorithme de détection de mouvement à cette multirésolution spatio-temporelle (section 6), on en améliore les performances dans le cas de la détection d'objets mobiles peu texturés ou d'objets lents (mouvement inférieur au pixel par image). La section 7 présente quelques résultats illustrant les avantages de la multirésolution avant de conclure dans la section 8.

2. rappels sur les champs de Markov

On adopte les notations suivantes :

- $s = (x, y, t)$ désigne le point de coordonnées (x, y, t) d'une tranche d'une séquence d'images;
- $I(x, y, t)$ représente la luminance du point s;
- $S = \{s \in [1, N_x] \times [1, N_y] \times [1, N_t]\}$ est l'ensemble des pixels ou sites d'une tranche (N_x, N_y et N_t désignent les dimensions spatio-temporelles de la tranche d'images) (cf. figure 1);
- $E = \{E_s, s \in S\}$ est le **champ des étiquettes cachées**;
- $e = \{e_s, s \in S\}$ est une réalisation particulière du champ E;
- $O = \{O_s, s \in S\}$ désigne le **champ des observations**;
- $o = \{o_s, s \in S\}$ est une réalisation particulière du champ O;
- e_s et o_s sont les valeurs des champs e et o au site s;
- Λ est l'ensemble des étiquettes possibles;

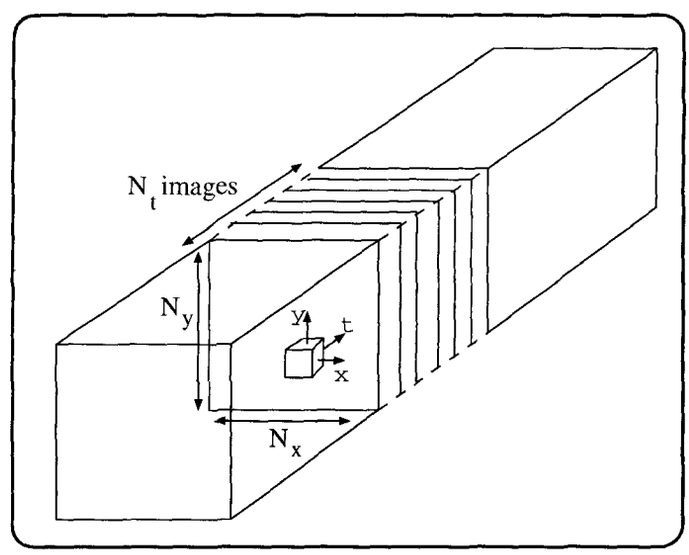

Figure 1. – Tranche de N_t images sur laquelle agit la relaxation spatio-temporelle.

– $R = \Lambda^S$ est l'ensemble des réalisations possibles du champ E.

Le problème à résoudre est de choisir parmi l'ensemble R des configurations possibles, la configuration e la plus probable de la séquence des étiquettes étant donné la séquence o des observations. Ce choix résulte de la relation suivante (critère du Maximum A Posteriori, MAP) :

$$\max_{e \in R} Pr[E = e/O = o] \Longleftrightarrow \max_{e \in R} Pr[E = e]\ Pr[O = oE = e] \quad (1)$$

où Pr désigne la probabilité. L'équivalence (1) résulte du théorème de Bayes en considérant que $Pr[O = o]$ est une constante vis-à-vis du problème traité.

On suppose que le champ E est un champ de Markov relativement à un voisinage $\mathcal{V}$ (cf. figure 2), c'est-à-dire que E vérifie les deux propriétés suivantes :

– $\forall e \in R\ \ Pr[E = e] > 0$

– $Pr[E_s = e_s\ /\ E_r = e_r, r \neq s, r \in S] = Pr[E_s = e_s\ /\ E_r = e_r, r \in \mathcal{V}]$

De la seconde propriété, il résulte que le choix (déterministe ou stochastique) de l'étiquette associée à un pixel s, conditionnellement à celles portées par tous les autres, ne dépend que des étiquettes portées par les seuls sites voisins [11]. E étant un champ de Markov, la maximisation du terme de probabilité conditionnelle $Pr[E = e/O = o]$ est équivalente à la **minimisation d'une fonction d'énergie globale** $U(e, o)$ constituée de deux termes [11] :

$$U(e, o) = U_m(e) + \lambda U_a(o, e) \quad (2)$$

– $U_m(e)$, **terme d'énergie associée au modèle a priori**, confère au modèle certaines propriétés pertinentes qui contraignent le problème et le rendent bien posé (régularisation);

– $U_a(o, e)$, **terme d'énergie d'attache aux données**, traduit l'accord de la solution avec les données observées.

– λ est une constante de pondération : elle permet de contrôler les influences respectives de chacun des deux termes d'énergie.

3. approche spatio-temporelle

Dans le modèle proposé, toutes les étiquettes des pixels d'une tranche d'images considérée seront estimées lors d'un même processus de relaxation. Nous parlerons donc de séquence d'étiquettes et de séquence d'observations sur une tranche plutôt que de champ d'étiquettes et champ d'observations. Dans la suite, même si ce n'est pas explicement spécifié, le terme séquence d'étiquettes devra être associé à une tranche temporelle limitée de la séquence d'images traitée (cf. figure 1).

3.1. étiquettes et observations pour la détection de mouvement

L'ensemble des étiquettes pertinentes associées à la détection de mouvement est $\Lambda = \{a, b\}$ où a est l'étiquette des points mobiles et b est l'étiquette des points fixes.

A la séquence d'images est associée une séquence d'étiquettes binaires E indiquant pour chaque pixel la présence ou l'absence de mouvement. On suppose que la séquence E des étiquettes « cachées » est un champ de Markov spatio-temporel. L'originalité de notre algorithme par rapport à [6, 13] concerne la définition d'un champ de Markov à trois dimensions (x, y, t) qui est par conséquent associé à un voisinage $\mathcal{V}$ cubique d'ordre 3 (cf. figure 2). La décision prise au point s va dépendre des étiquettes des 26 plus proches voisins de s. Il serait possible de considérer un voisinage plus étendu en espace et en temps, mais cela alourdirait beaucoup les calculs. On suppose ici que le voisinage cubique est pertinent, à savoir qu'il contient toutes les dépendances du pixel s et toutes les informations nécessaires à un choix correct de l'étiquette à affecter au pixel courant s.

On suppose que l'illumination est quasi constante entre deux images successives (variations lentes entre les instants t et $t + 1$) et que la caméra est fixe. Ces deux hypothèses permettent de considérer que toute variation temporelle significative de la fonction de luminance correspond à un mouvement dans la scène. Pour chaque point de la séquence, l'observation est donc définie par :

$$o_s = o(x, y, t) = |I(x, y, t + 1) - I(x, y, t)| \quad (3)$$

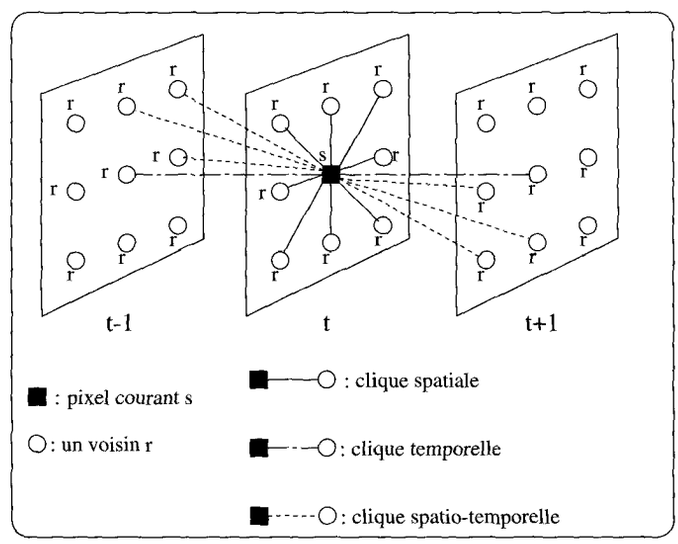

Figure 2. – Voisinage spatio-temporel $\mathcal{V}$ et cliques binaires $c = (r, s)$ associées (pour des raisons de clarté, toutes les cliques spatio-temporelles possibles n'ont pas été représentées).

3.2. énergie d'attache aux données

L'énergie d'attache aux données $U_a(o,e)$ est définie par :

$$U_a(o,e) = \frac{1}{2\sigma^2}\sum_{s\in S}[o_s - \Psi(e_s)]^2 \qquad (4)$$

où Ψ est définie par :

$$\Psi(e_s) = \begin{cases} 0 & \text{si } e_s = b \\ \alpha > 0 & \text{si } e_s = a \end{cases}$$

et σ^2 représente la variance des observations, qui peut être calculée sur chaque séquence d'observations. La fonction Ψ inspirée de celle proposée dans [6] modélise les observations. En effet, si le pixel appartient au fond fixe ($e_s = b$), il n'y a pas de changement temporel significatif de la fonction de luminance (l'illumination de la scène est supposée être lentement variable), donc l'observation est quasi nulle. En revanche, si le pixel appartient à un objet mobile ($e_s = a$), il y a changement temporel et on suppose que l'observation est proche d'une valeur α prédéfinie représentative de la valeur moyenne des observations non nulles. Il est possible d'estimer en ligne ce paramètre mais l'expérience montre que la qualité des résultats ne s'en trouve pas significativement améliorée [8].

3.3. énergie associée au modèle a priori

La forme générique de la fonction d'énergie est :

$$U_m(e) = \sum_{s\in S}\sum_{c\in C_s} V_c(e_s, e_r) \qquad (6)$$

où $V_c(e_s, e_r)$ est un potentiel élémentaire associé à la clique $c = (s,r)$ et où C_s est l'ensemble des cliques binaires du voisinage du pixel s considéré. L'objectif est d'obtenir des masques (projections sur le plan image des objets mobiles) homogènes en espace et en temps. On considère donc des fonctions de potentiel à niveau :

$$V_c(e_s, e_r) = \begin{cases} -\beta(s,r) & \text{si } e_s = e_r \\ +\beta(s,r) & \text{si } e_s \neq e_r \end{cases}$$

où le potentiel d'interaction $\beta(s,r)$ dépend de la nature de la clique $c = (s,r)$ considérée. En effet, sur le voisinage $\mathcal{V}$ de la figure 2, on distingue 5 types de cliques. Elles diffèrent les unes des autres par leur extension dans l'espace 3D contrairement à ce qui est fait dans [6, 13]. Si on note $\delta_x, \delta_y, \delta_t$ les coordonnées du vecteur représentatif de chaque clique dans l'espace 3D (x,y,t) centré en le pixel courant s, on définit (cf. figure 3) :

- 4 cliques spatiales horizontales et verticales (δ_x ou $\delta_y = \pm 1$, $\delta_t = 0$);
- 4 cliques spatiales diagonales (δ_x et $\delta_y = \pm 1$, $\delta_t = 0$);
- 2 cliques temporelles (δ_x et $\delta_y = 0$, $\delta_t = \pm 1$);
- 8 cliques spatio-temporelles horizontales et verticales (δ_x ou $\delta_y = \pm 1$, $\delta_t = \pm 1$);
- 8 cliques spatio-temporelles diagonales (δ_x et $\delta_y = \pm 1$, $\delta_t = \pm 1$).

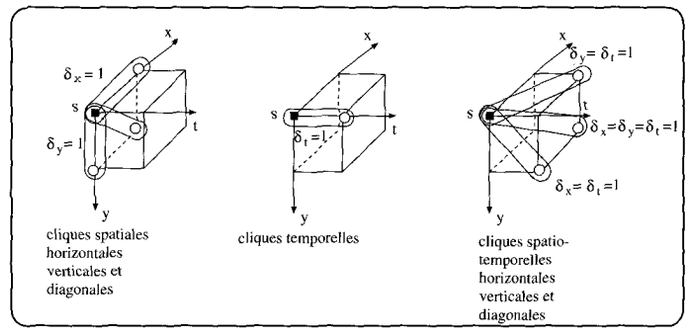

Figure 3. – Différents types de cliques binaires.

Il en résulte des cliques purement spatiales, purement temporelles ou spatio-temporelles. Il est alors naturel de définir un paramètre d'interaction spatiale β_s pour les dimensions spatiales (aucune distinction n'est faite entre les axes x et y) et un paramètre d'interaction temporelle β_t pour la dimension temporelle. Le paramètre $\beta(s,r)$ associé à chacune des cliques $c = (s,r)$ doit alors être une combinaison appropriée de ces deux paramètres de base. Puisque le potentiel $V_c(e_s, e_r)$ représente l'énergie d'interaction entre le pixel s et son voisin r, plus ces deux pixels sont éloignés, plus leur interaction mutuelle est faible. Il faut donc également faire intervenir dans la définition de $\beta(s,r)$ la distance euclidienne $d(s,r) = \sqrt{\delta_x^2 + \delta_y^2 + \delta_t^2}$ entre le pixel central s et le voisin r considéré. Ces remarques conduisent à la définition suivante :

$$\beta(s,r) = \frac{1}{d^2(s,r)\,(\frac{\delta_x^2}{\beta_s} + \frac{\delta_y^2}{\beta_s} + \frac{\delta_t^2}{\beta_t})} \qquad (8)$$

ce qui donne :

- $\beta(s,r) = \beta_s$ pour les cliques spatiales horizontales et verticales ($d(s,r) = 1$);
- $\beta(s,r) = \frac{\beta_s}{4}$ pour les cliques spatiales diagonales ($d(s,r) = \sqrt{2}$);
- $\beta(s,r) = \beta_t$ pour les cliques temporelles ($d(s,r) = 1$);
- $\beta(s,r) = \frac{\beta_s\beta_t}{2(\beta_s + \beta_t)}$ pour les cliques spatio-temporelles horizontales et verticales ($d(s,r) = \sqrt{2}$);
- $\beta(s,r) = \frac{\beta_s\beta_t}{3(\beta_s + 2\beta_t)}$ pour les cliques spatio-temporelles diagonales ($d(s,r) = \sqrt{3}$).

Cette fonction $\beta(s,r)$ ne fait intervenir que **deux paramètres** β_s **et** β_t. L'introduction de deux paramètres différents selon les axes spatiaux et temporel se justifie par le fait que l'espace et le temps sont des grandeurs de nature différente. Par ailleurs, dans la définition du terme d'énergie $U_a(o,e)$ (équation (4)), la dimension temporelle est favorisée par rapport aux dimensions spatiales étant donné la définition des observations qui sont des différences temporelles de la fonction de luminance. Afin de rééquilibrer les influences spatiale et temporelle dans l'expression de l'énergie globale (équation (2)), on choisit en pratique $\beta_s > \beta_t$.

Remarquons enfin que le calcul du paramètre d'interaction $\beta(s,r)$ ne dépend en fait que du type de clique considérée et non du pixel s considéré. Il est effectué une seule fois pour chacune des différentes cliques puis les résultats sont conservés dans une table de correspondance (*Look Up Table*) ce qui permet de ne pas avoir à recalculer $\beta(s,r)$ pour chaque point.

3.4. relaxation spatio-temporelle

Le critère du MAP se traduit en pratique par la recherche de la configuration e de la séquence des étiquettes qui minimise la fonction d'énergie $U(e,o)$. Cette minimisation est réalisée en utilisant l'algorithme de relaxation déterministe des ICM (*Iterated Conditional Modes*) [4]. Bien que cet algorithme itératif ne garantisse la convergence que vers le premier minimum local, il est utilisé car c'est le moins lourd en coût de calcul. Pour pallier le problème du piège des minima locaux, il est indispensable de soigner l'initialisation de la séquence des étiquettes. Celle-ci est initialisée par la séquence notée $\hat{E}$ des changements temporels binaires obtenue par binarisation de la séquence des observations grâce à une méthode statistique de maximum de vraisemblance qui s'appuie sur une modélisation linéaire de la fonction de luminance [2]. Cette méthode est plus robuste au bruit qu'un simple seuillage de l'observation en chaque point. Le calcul du champ des étiquettes initial nécessite le choix d'un seuil de binarisation θ. Ce choix est discuté par la suite.

Notre modèle de Markov étant spatio-temporel, il doit être associé à une relaxation spatio-temporelle. En effet, il est important que non seulement les voisins spatiaux mais aussi les voisins spatio-temporels et temporels soient remis à jour afin de propager dans le temps la contrainte markovienne d'homogénéité temporelle. Une version spatio-temporelle de l'algorithme des ICM a été mise en œuvre. Conformément à la figure 1, une itération de l'algorithme *spatio-temporel* de relaxation correspond à un balayage en x, y et t de la tranche de N_t images considérée. Les étiquettes de tous les pixels de cette tranche sont estimées. A la convergence, on obtient la séquence des étiquettes optimales sur cette tranche. La convergence est supposée atteinte lorsque la diminution relative de la fonction d'énergie globale sur une tranche d'images est inférieure à 10^{-5}. On note N_i le nombre d'itérations jusqu'à la convergence.

4. analyse critique des résultats

Lors de la description de l'algorithme spatio-temporel de détection de mouvement, la séquence d'images a été considérée comme un volume de données 3D. Pour des raisons évidentes de clarté, les résultats sont présentés sous la forme plus conventionnelle d'une succession d'images.

Afin de mettre en évidence la pertinence du modèle spatio-temporel, nous comparons ses performances à celles de l'algorithme de détection de mouvement construit sur la base d'un modèle markovien spatial décrit dans [8]. Ce modèle markovien de détection utilise la même expression pour la définition de $U_a(o,e)$ et une expression du même type pour la définition de $U_m(e)$ (fonctions de potentiel à niveau). Cependant, le voisinage considéré ne prend en compte que dix voisins (les huit plus proches voisins spatiaux et les deux plus proches voisins temporels). De plus, la relaxation a lieu à un instant donné (sur une image), les images étant traitées les unes après les autres. Sur les résultats présentés, une comparaison entre les masques obtenus grâce au modèle markovien spatial et ceux obtenus grâce au modèle markovien spatio-temporel est effectuée.

L'algorithme proposé requiert 4 paramètres pour le modèle $(\beta_s, \beta_t, \alpha, \lambda)$ et un seuil θ pour la construction de l'initialisation de la séquence des étiquettes. Il est aisé de fixer expérimentalement les valeurs de β_s, β_t, α et λ car une grande précision dans la détermination de ces valeurs n'est pas nécessaire. Il existe des méthodes pour estimer ces paramètres directement à partir des données [2], mais leur mise en œuvre n'a pas été envisagée afin de ne pas augmenter la charge de calculs. Nous avons pris $\beta_s = 20$, $\beta_t = 5$, $\alpha = 15$ et $\lambda = 5$ pour l'ensemble des séquences traitées. En revanche, le seuil θ de binarisation des observations, fixé lui aussi de manière expérimentale, doit être ajusté à la main pour chaque nouvelle séquence.

4.1. exemples

La figure 4 illustre l'efficacité de l'algorithme spatio-temporel pour traiter des séquences d'images bruitées. La séquence synthétique contient deux objets mobiles : un petit carré sombre qui se déplace de haut en bas à la vitesse de 2 pixels/image et un disque qui se translate vers la droite à la même vitesse. Un bruit impulsionnel non corrélé en espace et en temps est introduit sur les niveaux de gris moyens de chaque image. Le disque ayant un mouvement assez lent (par rapport à sa taille) d'une image à l'autre, l'intersection entre les positions successives de cet objet est non vide et il est nécessaire de prendre un seuil de binarisation faible afin d'obtenir une séquence d'étiquettes initiales significative pour la zone de glissement de l'objet sur lui-même. Il en résulte un bruit important dans les zones statiques sur cette

séquence d'étiquettes initiales (cf. ligne 2 de la figure 4). Sur les masques obtenus après relaxation spatiale, on relève un nombre important de fausses détections (cf. ligne 3 de la figure 4). En revanche, la relaxation spatio-temporelle permet l'élimination complète de ces fausses détections (cf. ligne 4 de la figure 4). On remarque que le masque associé au disque est un peu gros, ce qui est à mettre en relation avec la taille du voisinage spatio-temporel utilisé dans le modèle et avec le fait qu'aucune information de contour n'est prise en compte.

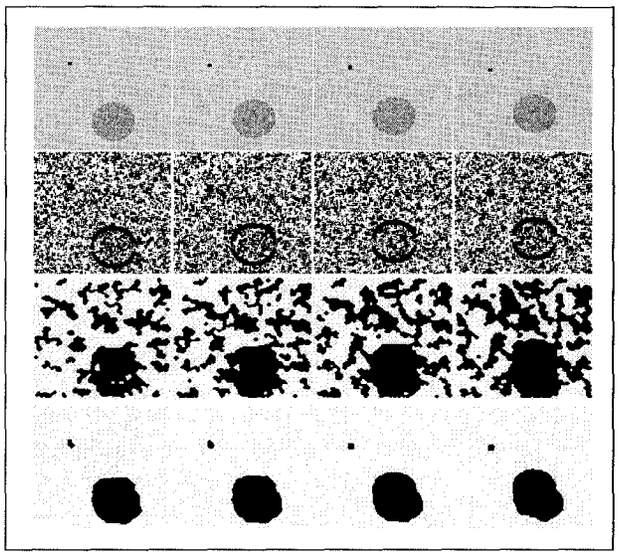

Figure 4. – De haut en bas : 1) séquence synthétique; 2) initialisation binaire de la séquence des étiquettes; 3) masques finaux après relaxation *spatiale*; 4) masques finaux après relaxation *spatio-temporelle* ($\theta = 20$, $N_t = 8$, $N_i = 7$).

L'autre avantage de l'algorithme *spatio-temporel* réside dans sa capacité à reconstruire les masques d'objets dont le mouvement a lieu avec glissement (intersection non vide des masques entre deux instants consécutifs). Dans la zone de glissement des objets sur eux-même, l'information de mouvement est très pauvre. La caractéristique de l'algorithme spatio-temporel est qu'il permet la propagation d'information en espace et en temps. La séquence d'images de la figure 5 contient deux zones mobiles : un groupe de piétons marchant sur le trottoir (deux piétons sombres se déplaçant vers la gauche et un piéton plus clair se déplaçant vers la droite) et un vélo se déplaçant vers la gauche. Après la première itération, sur la première image, le masque relatif aux piétons n'est que partiellement reconstruit du fait du manque d'information de mouvement dans cette zone de glissement. Mais, sur la troisième image, il y a assez d'information de mouvement dans la zone de glissement pour reconstruire entièrement le masque des piétons. Grâce au balayage spatio-temporel, il est alors possible de propager dans le temps les contraintes markoviennes d'homogénéité spatiale et temporelle des masques et donc de retrouver entièrement le masque des piétons de la première image après quelques itérations (en l'occurrence, à la troisième itération). Au contraire, l'agorithme à relaxation *spatiale* induit un processus causal et ne permet pas de revenir dans le temps sur une décision. Les itérations supplémentaires jusqu'à la convergence servent à éliminer la zone d'écho (zone de fond découverte lors du mouvement), en particulier pour le vélo (comparer les lignes 3 et 4 de la figure 5).

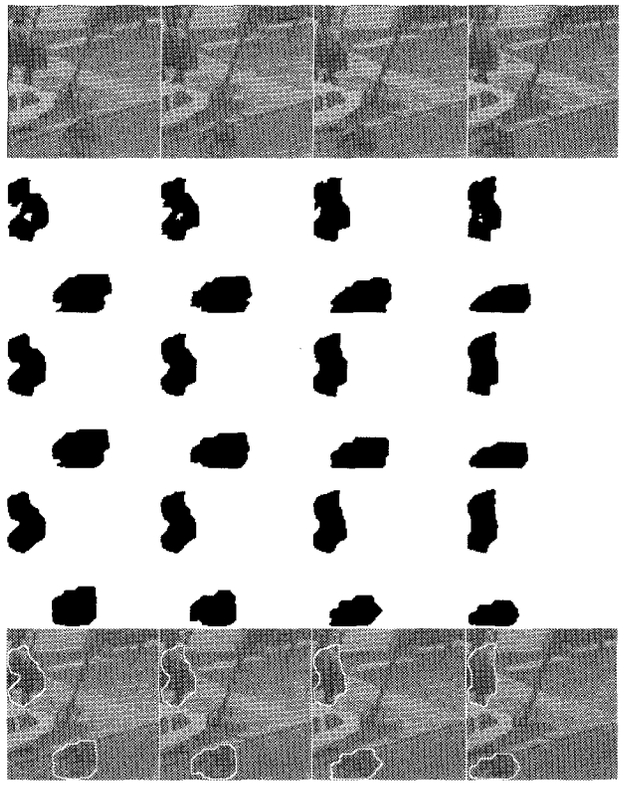

Figure 5. – De haut en bas : 1) scène de rue; 2) masques pour $N_i = 1$; 3) masques pour $N_i = 3$; 4) masques finaux pour $N_i = 15$ ($\theta = 32$, $N_t = 8$); 5) superposition des contours des masques sur les images originales.

4.2. choix de N_t

Ce sont les choix du voisinage et des fonctions de potentiel qui définissent totalement le modèle markovien. La valeur de N_t n'a pas d'incidence sur ces deux caractéristiques du modèle. Donc le choix de N_t n'a pas de conséquence sur la modélisation proprement dite. En pratique, N_t contrôle la portée de la contrainte markovienne d'homogénéité temporelle par l'intermédiaire de la phase de relaxation. Concernant l'étendue de cette contrainte d'homogénéité temporelle, le choix de N_t n'est pas évident. Plusieurs valeurs de N_t ont été testées afin d'évaluer précisément l'influence de ce paramètre [8]. Les cas pour lesquels la valeur de N_t peut avoir une incidence (par exemple, obtention de masques trop gros si N_t est trop grand) sont ceux où la contrainte d'homogénéité spatio-temporelle des masques est rompue, i.e. lors de l'apparition ou la disparition d'objets ou encore lors de la séparation d'objets connexes. Le choix de N_t est également important lors du traitement de séquences très bruitées. Dans un

tel cas, on a intérêt à augmenter la valeur de N_t afin d'éliminer toutes les détections parasites.

Par ailleurs, le choix de N_t a évidemment une incidence sur la quantité de mémoire nécessaire pour la mise en œuvre ainsi que sur le retard d'obtention des résultats. Les mises en œuvre sont d'autant moins gourmandes en mémoire et temps de calcul (donc plus rapides) que le flux des données à analyser est moins important. Il est également préférable que le retard soit le plus faible possible. Ces contraintes conduisent au choix d'une valeur de N_t la plus faible possible. La valeur minimale admissible est $N_t = 5$ à cause des effets de bord temporels : étant donné le voisinage choisi (cf. figure 2), les première et dernière images de chaque séquence ne sont pas traitées par la relaxation.

En revanche, puisque l'ensemble des masques de la tranche temporelle N_t est obtenu en même temps, il en résulte un retard compris entre 1 et N_t images, soit en moyenne un retard de $\frac{N_t}{2}$ images.

D'après notre expérience, pour une séquence quelconque, la valeur $N_t = 8$ est préconisée par défaut, sachant qu'il est éventuellement possible de rectifier les résultats en ajustant ce paramètre (par exemple, augmenter N_t si la séquence est très bruitée, diminuer N_t si les mouvements sont rapides ...)

4.3. complexité calculatoire

Le temps *cpu* moyen de traitement d'une image de taille 128×128 sur une station de travail Sun Sparc-10 pour un algorithme programmé en langage C non optimisé est de $4s$. Ce temps de traitement est incompatible avec des contraintes de temps réel. Pour pouvoir utiliser cet algorithme dans le cadre d'une application, il faut obligatoirement étudier sa mise en œuvre sur du matériel spécifique. A titre de solution possible, une implantation sur une machine parallèle non dédiée a été effectuée [8]. La valeur de N_t a été fixée à 5 afin de limiter le flux de données traitées. La cadence de traitement atteinte est de 3 à 4 images/seconde ce qui est encourageant pour un algorithme qui semble a priori lourd en calculs. L'implantation sur une machine dédiée permettrait d'augmenter cette cadence et d'envisager l'utilisation de cet algorithme dans le cadre d'applications telles que le contrôle du trafic routier ou la surveillance de sites.

5. approche multirésolution spatio-temporelle

En cas de mouvement très lent (déplacement inférieur au pixel par image) ou de mouvement d'objets d'intensité uniforme, l'observation (équation (3)) est très pauvre bien qu'un mouvement soit effectivement présent. De ce fait, les performances de l'algorithme ne sont pas bonnes (détection partielle des objets mobiles). Pour renforcer l'information de mouvement, une solution consiste à construire l'observation à partir d'un horizon spatial et temporel plus étendu. On estime successivement le champ de Markov sur chaque niveau d'une pyramide construite à la suite d'une série de filtrages et de sous-échantillonnages. L'approche multirésolution se compose de trois étapes :

- construction de la pyramide;
- stratégie de parcours de la pyramide;
- adaptation du modèle markovien à chaque niveau.

5.1. construction de la pyramide

Il existe de nombreux types de pyramides qui diffèrent essentiellement par la nature et l'agencement des filtres utilisés. On distingue les transformations pyramidales passe-bas, passe-bandes orthogonales ou non, les pyramides d'ondelettes ou quinconces [3, 7, 14]. Etudions quelle est la transformation pyramidale la plus appropriée vis-à-vis du modèle de détection. L'intégration du modèle de détection de mouvement dans un cadre multirésolution est motivée par le désir d'améliorer la qualité des résultats dans deux situations précises : mouvement sous-pixel et mouvement d'objets de luminance uniforme. Dans cette optique, une transformation pyramidale passe-bas semble la plus adaptée car elle permet de moyenner les informations sur un domaine spatial ou temporel plus large. Ayant défini un modèle de Markov 3D pour modéliser la séquence des masques, on se propose de construire non pas une séquence de pyramides d'images (multirésolution spatiale uniquement, cf. figure 6-a) mais bien *une pyramide de séquences d'images* (cf. figure 6-b). Il est classique de créer une pyramide spatiale d'images pour traiter le problème des objets peu texturés. Nous y ajoutons une pyramide temporelle d'images afin de traiter le cas des mouvements sous-pixels.

La pyramide de séquences est construite par extension au cas 3D de la pyramide de Burt, i.e. par une succession de filtrages et sous-échantillonnages dont la caractéristique est qu'ils sont spatio-temporels au lieu d'être uniquement spatiaux (cf. figure 7-a). Le noyau du filtre passe-bas est un noyau tridimensionnel (cf. figure 7-b) construit à partir du filtre 1D élémentaire $\frac{1}{4}[1\ 2\ 1]$ appliqué dans les trois directions (x, y, t). L'opération de filtrage est ainsi la même dans les trois directions. Ayant limité par ce filtrage la bande des fréquences spatiales et temporelles, un sous-échantillonnage sans repliement de spectre est possible et réduit la taille de la séquence d'un facteur 2 dans chacune des trois directions (la séquence au niveau de résolution inférieure comprend deux fois moins d'images de taille quatre fois plus petite). Un exemple de pyramide spatio-temporelle est donné sur la figure 8 pour trois niveaux ($k = 0$ correspond à la résolution la plus fine et $k = 2$ à la résolution la plus grossière).

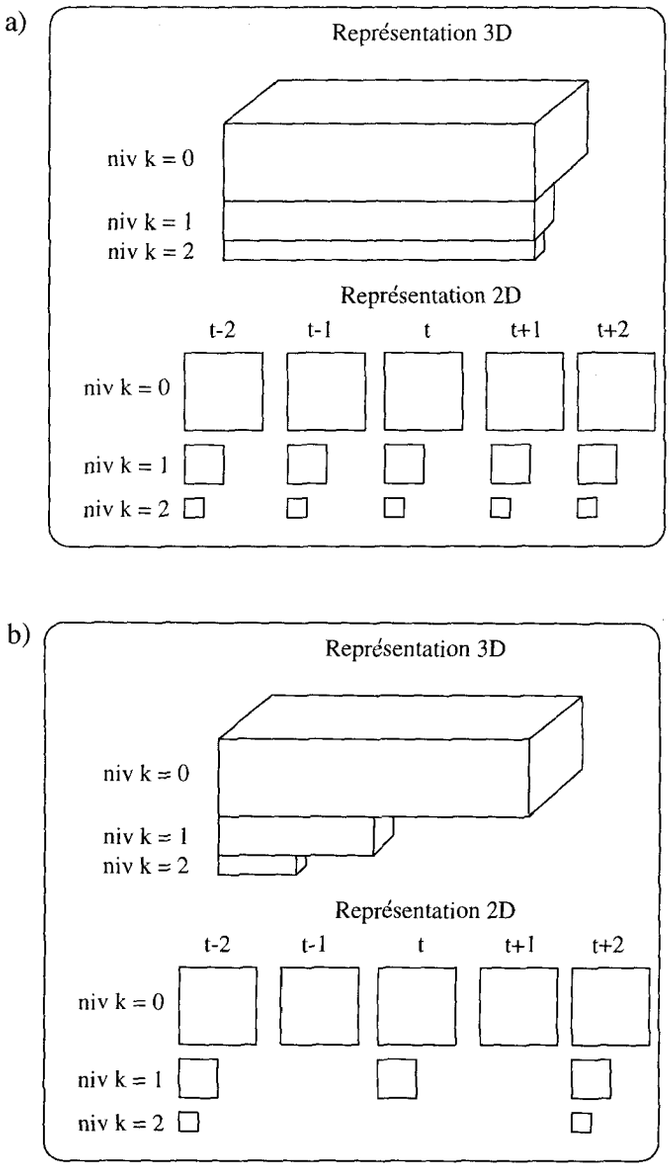

Figure 6. – a) Multirésolution spatiale (séquence de pyramides. b) Multi-résolution spatio-temporelle (pyramide de séquence).

Pour construire une image de la pyramide au niveau 2, il faut disposer de 7 images consécutives de la séquence initiale (cf. figure 9). Soit une portion de la séquence (par exemple, une vingtaine d'images) dont on construit la pyramide. Si cette pyramide est à 3 niveaux, il y a 10 images au niveau 1 et 5 images au niveau 2. Toutes les images d'un niveau sont traitées en même temps ($N_t = 5$ pour $k = 2$, $N_t = 10$ pour $k = 1$ et $N_t = 20$ pour $k = 0$). Cela revient à augmenter d'un facteur 2 la taille de la section temporelle considérée au niveau de résolution plus fine de la pyramide.

La séquence est découpée en tranches d'images qui sont explorées successivement. Le nombre d'images à prendre en compte doit être tel qu'au niveau le plus bas, on ait une séquence d'au moins

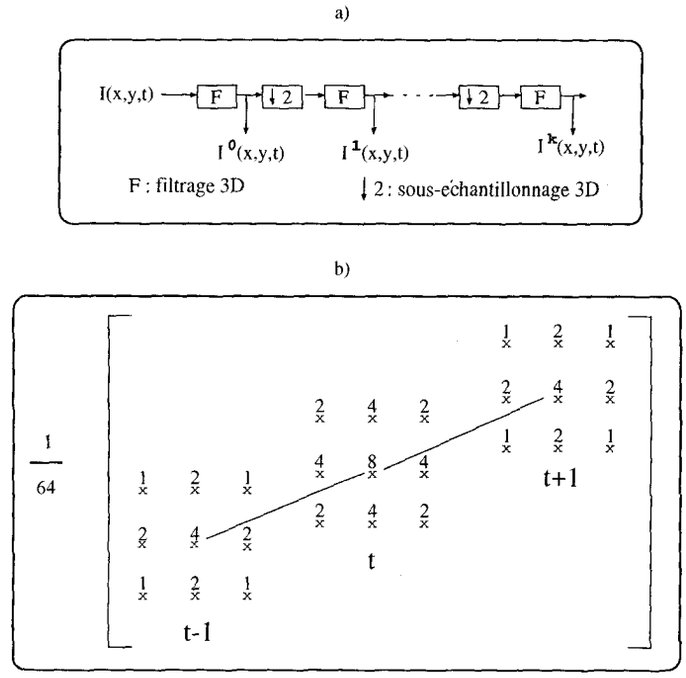

Figure 7. – a) Principe de construction de la pyramide. b) Noyau du filtre spatio-temporel 3D.

5 images (valeur minimale admissible à cause des effets de bord temporels cf. paragraphe 4.2).

5.2. parcours de la pyramide

La stratégie de parcours de la pyramide spatio-temporelle est classique, il s'agit de l'approche descendante ou *coarse to fine* : la relaxation débute sur le niveau le plus grossier, le résultat est interpolé et sert d'initialisation au niveau de résolution immédiatement supérieure. Ce processus est répété jusqu'au niveau de résolution la plus fine. La spécificité de notre approche est que l'interpolation est elle aussi spatio-temporelle (cf. figure 10) : l'étiquette d'un point s au niveau k sert d'initialisation pour 8 points au niveau $k-1$ (rééchantillonnage d'un facteur 2 dans chacune des directions x, y, et t).

Avec un tel parcours, on s'affranchit du problème de convergence vers un minimum local. En effet, il a été constaté que la fonction d'énergie est plus lisse (moins de minima locaux) à des niveaux de résolution plus grossière. Donc dès le niveau de résolution la plus grossière, on se dirige vers un meilleur minimum local (si ce n'est vers le minimum global). En prenant comme champ d'étiquettes initial au niveau $k-1$ le résultat interpolé de la relaxation au niveau k, on se trouve proche du minimum global dès la première itération.

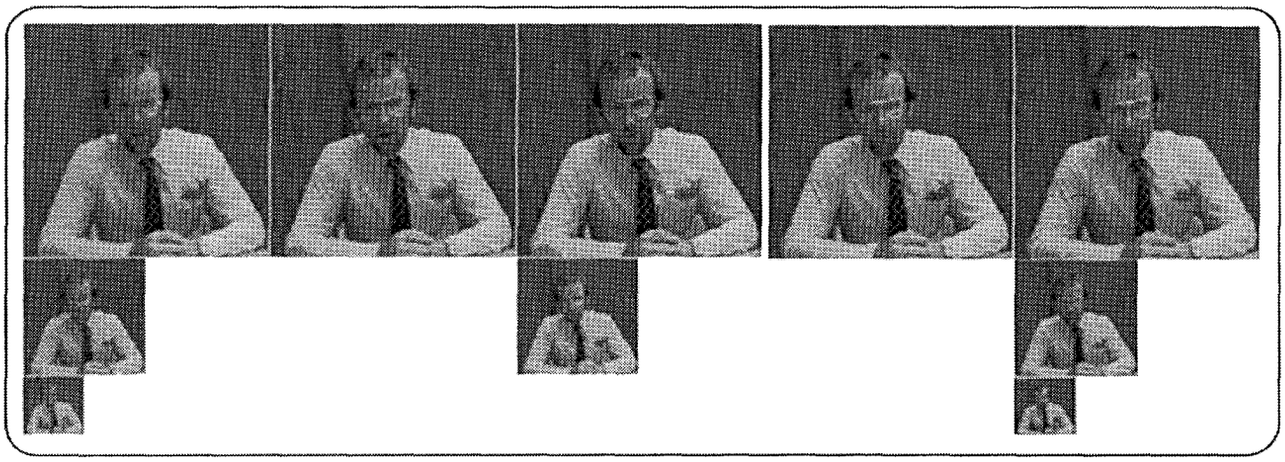

Figure 8. – Exemple de pyramide spatio-temporelle à trois niveaux sur la séquence *Trevor* (de haut en bas : $k = 0, 1, 2$).

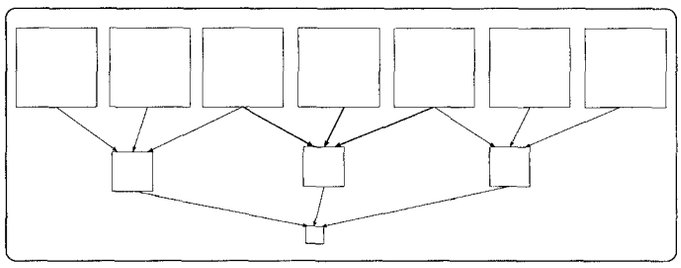

Figure 9. – Construction d'une image de la pyramide au niveau $k = 2$.

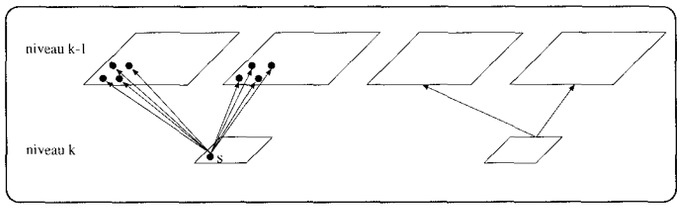

Figure 10. – Interpolation spatio-temporelle pour l'initialisation de la séquence des étiquettes.

6. détection de mouvement multirésolution

Contrairement à d'autres approches hiérarchiques telles que l'approche multiéchelle par exemple [16], la difficulté majeure en analyse multirésolution est l'adaptation du modèle markovien à chaque niveau de la pyramide (évolution des voisinages, des cliques et des fonctions de potentiel). Dans la majorité des cas, soit le même modèle est conservé à chaque niveau, soit les évolutions sont empiriques. Dans ce qui suit, on s'efforce de justifier au mieux les lois d'évolution imposées aux différents paramètres du modèle.

6.1. synoptique de l'algorithme

La figure 11 détaille le flux des données dans chaque phase de l'algorithme pour une pyramide à trois niveaux. A partir de la séquence d'images, on construit la séquence O des observations. A priori, il y a deux stratégies possibles pour construire les observations à chacun des niveaux de la pyramide : soit on construit la pyramide des données et on calcule les différences inter-images à chaque niveau de cette pyramide; soit on calcule les observations pleine échelle à partir de la séquence d'images initiale puis on élabore la pyramide des observations. Ces deux techniques de construction des observations multirésolution ne sont pas identiques car le calcul des observations n'est pas une opération linéaire (cf. valeur absolue). Expérimentalement, la seconde solution est apparue être la plus satisfaisante. De ce fait, conformément au diagramme de la figure 7a, le premier niveau des observations O^0 est égal aux observations pleine échelle après un premier filtrage 3D. La séquence des étiquettes est ensuite estimée successivement à chaque niveau étant donné la séquence des observations correspondante (cf. figure 11). La figure 12 donne le synoptique de l'algorithme multirésolution pour une pyramide à trois niveaux ($k = 0, 1, 2$). Remarquons qu'au niveau de résolution la plus grossière ($k = 2$ dans le cas de la figure 12), la séquence des étiquettes est initialisée par une séquence d'étiquettes binaires $\hat{E}^k$ issue de la binarisation par test de maximum de vraisemblance des observations au niveau de résolution la plus grossière. Pour les autres niveaux, l'initialisation $\hat{E}^{k-1}$ découle de l'interpolation (cf. figure 10) des résultats de relaxation E^k du niveau de résolution inférieure.

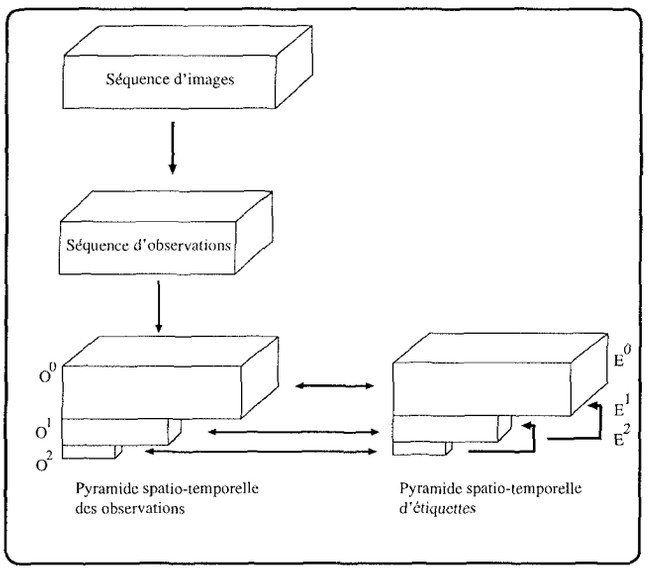

Figure 11. – Etiquettes et observations associées.

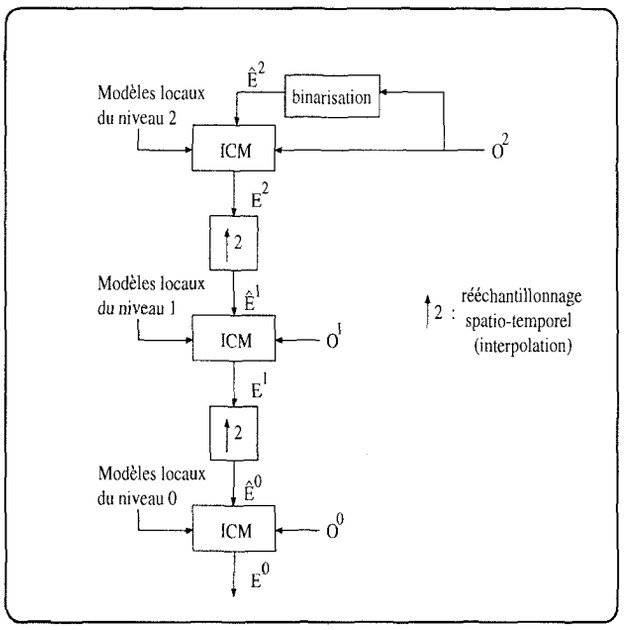

Figure 12. – Synoptique de l'algorithme multirésolution.

6.2. évolution des observations

Contrairement à l'approche multiéchelle pour laquelle un seul niveau d'observations (pleine échelle) est utilisé [16], dans l'approche multirésolution, une pyramide spatio-temporelle d'observations est construite pour obtenir la pyramide spatio-temporelle des étiquettes (cf. figure 11). Etudions l'évolution des observations le long de la pyramide.

Atténuation de l'amplitude. La pyramide des observations est obtenue suite à une succession de filtrages passe-bas. Il en résulte une élimination du bruit ainsi qu'une atténuation de l'amplitude des variations temporelles. Considérons le déplacement d'un front vertical à la vitesse de 1 pixel/image (cf. figure 13). A la suite du filtrage passe-bas spatio-temporel, l'amplitude de l'observation liée au déplacement du front est réduite d'un facteur $\frac{8}{3} = 2.6$. A chaque niveau successif, l'amplitude des observations est donc réduite d'un tiers environ.

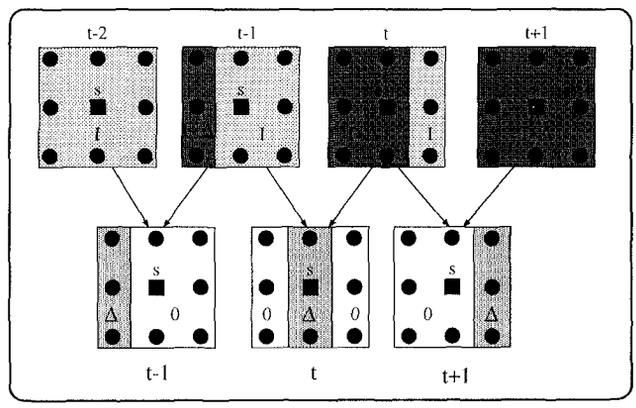

Figure 13. – Évolution des observations lors du filtrage spatio-temporel : cas du déplacement d'un front vertical. Avant filtrage 3D : $I(x,y,t-1) = I$, $I(x,y,t) = I + \Delta$, $o(x,y,t) = |I(x,y,t) - I(x,y,t-1)| = |\Delta|$. **Après filtrage 3D :** $o(x,y,t) = \frac{1}{64}(4|\Delta| + 16|\Delta| + 4|\Delta|) = \frac{3}{8}|\Delta|$

Amélioration de la qualité des observations. La séquence synthétique de la figure 14 contient trois objets mobiles : un rectangle clair qui se déplace vers la droite à la vitesse de 1 pixel/image, un carré noir qui se déplace vers la gauche à la même vitesse et, en bas à gauche, un second carré sombre dont le mouvement est une translation verticale descendante à la vitesse de 0.35 pixel/image. Les observations sont données d'une part en monorésolution et d'autre part pour deux niveaux ($k = 0, 1$) en multirésolution spatio-temporelle. On constate que pour le carré le plus à gauche, au mouvement sous-pixel, des variations temporelles sont enregistrées pour chaque image en multirésolution alors qu'elles n'apparaissent que toutes les deux images en monorésolution. Par ailleurs, pour le rectangle clair peu contrasté par rapport au fond et de déplacement faible par rapport à sa taille, les observations sont également plus marquées dans la zone de glissement de l'objet sur lui-même pour le cas multirésolution. La figure 14 présente aussi les observations pour la séquence *Trevor* avec trois niveaux ($k = 0, 1, 2$). Cette séquence représente le mouvement d'un présentateur de télévision : le mouvement entre deux images successives est faible et certaines zones de la chemise ou encore des mains sont peu texturées. Les mêmes remarques que pour la séquence synthétique peuvent être faites vis-à-vis de la comparaison des observations monorésolution et multirésolution.

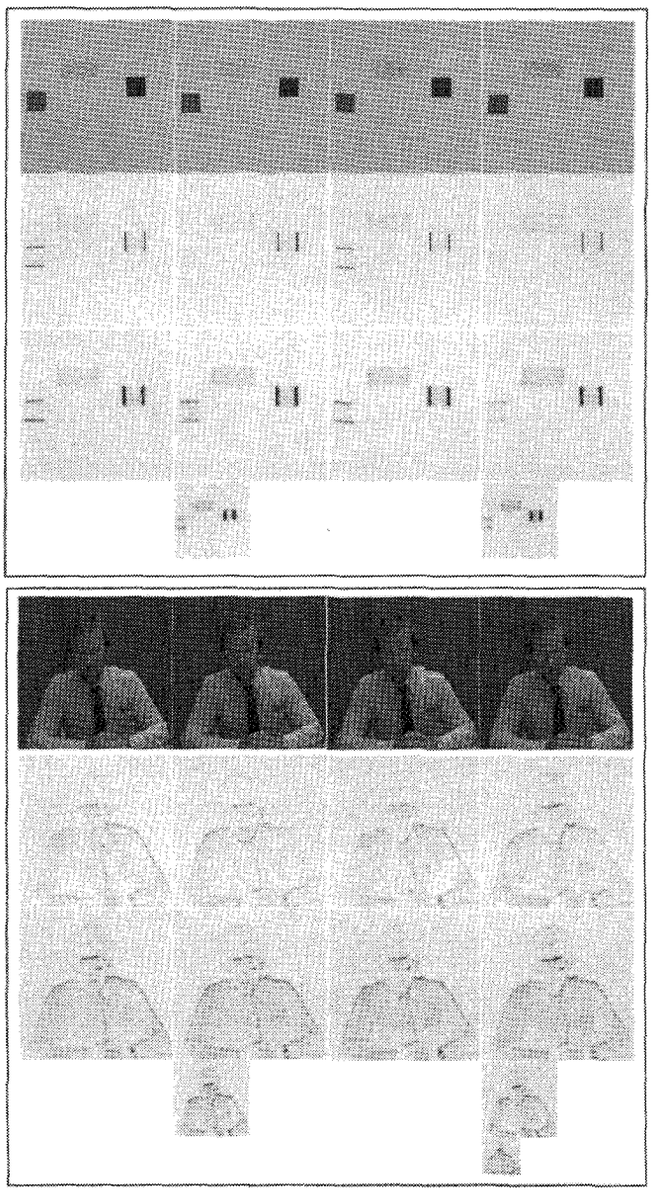

Figure 14. – Évolution des observations le long de la pyramide spatio-temporelle pour deux séquences différentes (une séquence synthétique et la séquence Trevor) : de haut en bas : 1) séquence d'images; 2) observations monorésolution (plus le pixel est sombre, plus l'observation associée est de grande amplitude.); 3) observations multirésolution, 2 ou 3 niveaux d'analyse ($k = 0, 1$ ou 2).

En conclusion, le filtrage spatio-temporel améliore les observations dans les deux cas (mouvement très lent et mouvement d'objets peu texturés) où le modèle de détection monorésolution est mis en défaut justement à cause de la pauvreté des observations. Ceci vient du fait que la multirésolution proposée intègre l'information sur un horizon spatial et temporel plus large. En effet, une image de la séquence O^0 intègre l'information de mouvement contenue dans trois images consécutives de la séquence initiale des observations (un seul filtrage) alors qu'une image au niveau $k = 2$ intègre l'information de mouvement contenue dans 7 images successives de la séquence des observations initiales (3 filtrages). Ne considérer qu'un seul filtrage revient à appliquer l'algorithme sur une pyramide à un seul niveau de résolution, ce qui est insuffisant pour les exemples présentés (masques partiellement reconstruits). En revanche, l'approche multirésolution est déconseillée pour détecter le mouvement d'objets rapides car le filtrage temporel conduit à des observations lissées de sorte que les masques reconstruits sont beaucoup plus gros que les masques réels (perte accrue de précision sur les frontières de mouvement à cause du filtrage temporel qui effectue une moyenne dans le temps de tous les mouvements présents).

6.3. évolution de l'énergie d'attache aux données

La fonction Ψ modélisant le lien entre étiquettes et observations fait intervenir le paramètre α, constante représentative de l'amplitude moyenne des observations non nulles. Nous avons vu que cette amplitude est réduite d'un tiers à la suite de chaque filtrage donc ce paramètre doit évoluer en conséquence :

$$\alpha_k = \frac{1}{3}\alpha_{k-1} \tag{9}$$

Il résulte de cette évolution une réduction d'un facteur 9 du terme $(o_s - \Psi(e_s))^2$ qui est compensée par la diminution dans un même rapport de la variance σ^2 des observations. L'énergie d'attache aux données est donc quasi constante le long de la pyramide. D'autre part, on a montré l'amélioration de la qualité des observations le long de la pyramide (cf. figure 14). Il est pertinent d'augmenter l'influence du terme d'attache aux données (plus fiables) par rapport au terme de régularisation (cf. paragraphe 6.5).

6.4. évolution de l'énergie associée au modèle a priori

$U_m(e)$ modélise les interactions spatiales et temporelles entre pixels. Puisque le filtrage servant à la construction de la pyramide est le même dans les trois directions, les paramètres β_s et β_t du modèle doivent être modifiés dans les mêmes proportions. En effet, interactions spatiales et temporelles ont été modifiées de la même manière. En revanche, il est logique de penser que ces interactions faiblissent lorsque la résolution devient plus grossière puisque deux pixels voisins à un niveau de résolution grossière sont en réalité beaucoup plus éloignés au niveau de résolution la plus fine. Pour traduire ce phénomène, il faut réduire les coefficients β_s et β_t le long de la pyramide.

6.5. loi d'évolution proposée

Tenant compte de 6.3 et 6.4 et puisque l'algorithme de relaxation repose sur un équilibre entre les deux termes d'énergie, nous proposons de regrouper dans une même loi les phénomènes de diminution de $U_m(e)$ et d'augmentation de $U_a(o,e)$. La loi d'évolution, choisie de façon expérimentale, consiste à pondérer d'un facteur :

$$\lambda_k = 4^k \lambda \tag{10}$$

l'énergie d'adéquation au niveau k, l'expression de l'énergie totale étant :

$$U(e,o) = U_m(e) + \lambda_k U_a(o,e) \tag{11}$$

La réduction de $U_m(e)$ est bien intégrée en pratique dans la loi d'évolution de $U_a(o,e)$ car il est équivalent de diminuer $U_m(e)$ et d'augmenter $U_a(o,e)$.

7. résultats

La version multirésolution spatio-temporelle de l'algorithme de détection de mouvement a été mise en œuvre et testée sur un ensemble de séquences synthétiques et du monde réel.

7.1. mouvement sous-pixel

La multirésolution améliore avant tout les performances de l'algorithme pour détecter des objets dont le mouvement est inférieur au pixel par image. En effet, dans un tel cas, l'observation est très pauvre. Sur la figure 15, on présente les masques obtenus en monorésolution et en multirésolution pour deux niveaux d'analyse. En monorésolution, le masque du carré au mouvement sous-pixel est quasi inexistant, celui du rectangle est partiellement reconstruit. En revanche, en multirésolution, les masques des trois objets mobiles sont entièrement reconstruits dès le niveau de résolution la plus grossière, le filtrage temporel sur plusieurs images successives de la séquence renforçant l'information de mouvement disponible.

7.2. objets d'intensité uniforme

L'autre cas intéressant pour l'utilisation de l'algorithme multirésolution est la détection des objets mobiles peu texturés. Le filtrage spatial permet de renforcer la quantité d'information présente dans les observations. Sur la figure 16, on présente les masques monorésolution et multirésolution (3 niveaux d'analyse) pour une séquence d'images comprenant deux piétons en mouvement, celui de droite étant particulièrement peu texturé au niveau du manteau. Les masques obtenus en monorésolution sont incomplètement reconstruits pour le piéton de droite. Au contraire,

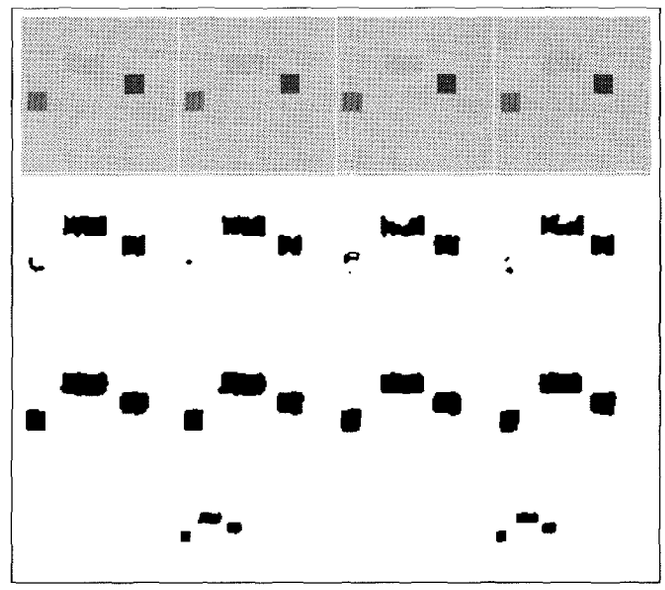

Figure 15. – Détection de mouvement sous-pixel : de haut en bas : 1) séquence synthétique; 2) masques monorésolution; 3) masques multirésolution (2 niveaux d'analyse : $k = 0, 1$).

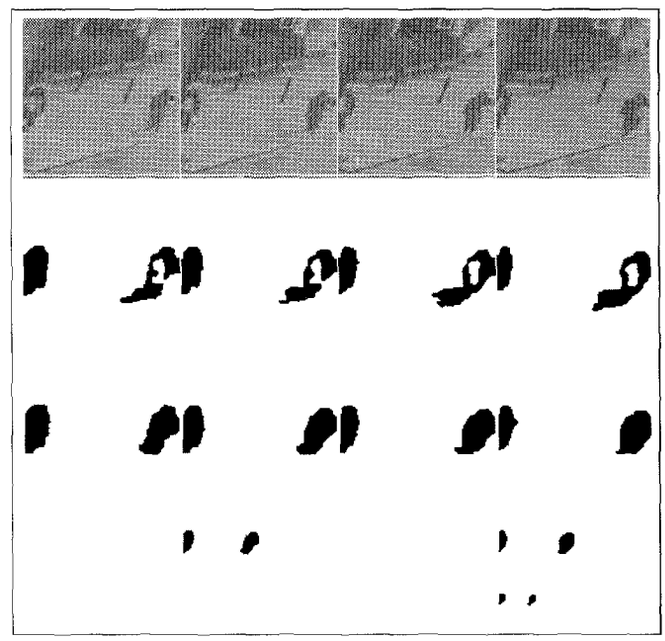

Figure 16. – Détection de mouvement d'objets peu texturés : de haut en bas : 1) séquence de scène de rue; 2) masques monorésolution; 3) masques multirésolution (3 niveaux d'analyse : $k = 0, 1, 2$).

ces masques sont complets dès le niveau de résolution la plus basse avec la version multirésolution de l'algorithme.

La séquence de la figure 17 est un exemple qui typiquement doit être traité par l'algorithme multirésolution. En effet, le mouvement du présentateur est lent d'une image à l'autre et le personnage contient des zones uniformes en mouvement. Les masques monorésolution sont très incomplets tant au niveau de la tête dont l'amplitude du mouvement est faible qu'au niveau

du bas de la chemise qui représente une zone peu texturée. La multirésolution permet de s'affranchir de ces difficultés.

7.3. complexité calculatoire

L'utilisation d'un cadre multirésolution conjointement à la modélisation markovienne est généralement motivée par le désir d'accélérer les temps de traitement, l'idée étant que des calculs locaux sur une grille grossière d'une région donnée sont équivalents à des calculs plus globaux sur une grille plus fine couvrant la même région. Par ailleurs, lorsque l'algorithme de relaxation est un algorithme déterministe, donc sous-optimal, la multirésolution permet d'améliorer la phase d'initialisation. Pour évaluer la réduction du coût de calcul, un nombre d'itérations équivalent noté N_{ieq} jusqu'à la convergence dans le cas multirésolution a été estimé, sachant qu'une itération à pleine échelle correspond à $\frac{1}{2^{3k}}$ itérations au niveau k (2^k fois moins d'images de taille $2^k \times 2^k$ fois plus petite). Sur le tableau 1, N_{ieq} a été évalué pour diverses séquences (une image de chacune des séquences traitée dans ce tableau est donnée sur la figure 18). L'examen de ces résultats montre que l'accélération des calculs n'est pas la caractéristique principale de l'algorithme multirésolution ce qui s'explique par le fait que la relaxation monorésolution demande déjà peu d'itérations. Le rôle principal de la multirésolution est l'amélioration des observations dans les cas critiques mentionnés précédemment.

Tableau 1. – Comparaison du nombre d'itération jusqu'à la convergence en monorésolution et en multirésolution.

séquence	1	2	3	4
monorésolution N_i	5	7	11	6
multirésolution $\|\| N_i$ *niv*2	3		3	3
multirésolution $\|\| N_i$ *niv*1	4	3	4	4
multirésolution $\|\| N_i$ *niv*0	4	4	2	5
multirésolution $\|\| N_{ieq}$	4.55	4.4	2.56	5.6

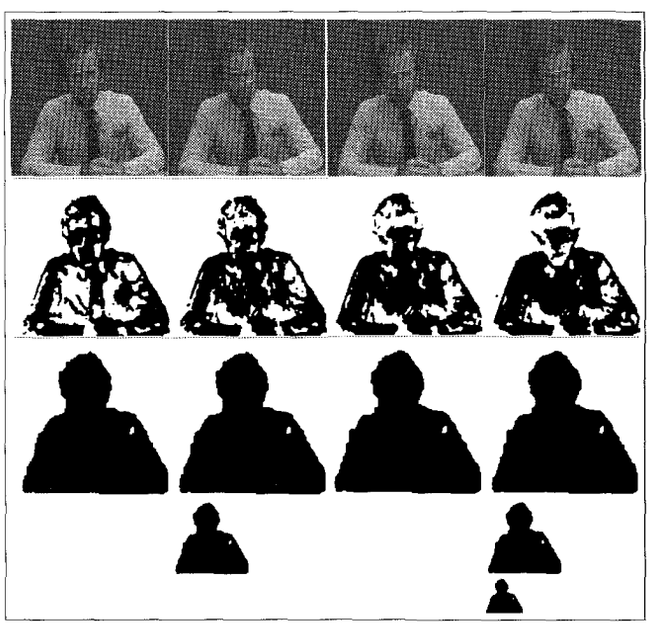

Figure 17. – Détection de mouvement multirésolution : 1) séquence Trevor; 2) masques monorésolution; 3) masques multirésolution (3 niveaux d'analyse : $k = 0, 1, 2$).

8. conclusion

Classiquement, les algorithmes d'analyse du mouvement dans les séquences d'images font intervenir deux ou trois images successives. La stratégie proposée ici consiste à considérer une séquence d'images comme un flux de données 3D. Ceci permet d'augmenter la quantité d'informations prises en compte pour prendre une décision en un point donné. Dans ce cadre, un modèle markovien spatio-temporel associé à une relaxation elle-même spatio-temporelle a été introduit. En s'appuyant uniquement sur l'information de mouvement assez rudimentaire que constitue la différence entre deux images successives, mais en modélisant finement les interactions spatio-temporelles dans un voisinage cubique autour de chaque pixel, l'algorithme proposé est plus performant que son analogue 2D. La fiabilité de détection d'objets mobiles dans des séquences très bruitées est augmentée et la reconstruction des zones de glissement des objets sur eux-mêmes est améliorée. L'inconvénient de l'intégration temporelle sur un plus grand nombre d'images est que le délai d'obtention des résultats est accru ce qui peut être gênant dans certains types d'applications temps réel. En définissant ce modèle 3D du champ d'étiquettes, on réussit à exploiter au mieux les informations de mouvement contenues dans la simple différence temporelle entre deux images. Néanmoins, cette observation s'avérant trop rudimentaire pour traiter le cas des objets au déplacement très lent et celui des objets très peu texturés, elle est améliorée par la construction d'une pyramide spatio-temporelle de séquences d'images. Les résultats présentés témoignent du bien fondé de cette approche. Cependant, la multirésolution est à déconseiller dans le cas de mouvements rapides, le filtrage temporel engendrant une dégradation des résultats par rapport à la monorésolution.

Avec les deux versions (mono et multirésolution) de l'algorithme, il est possible de traiter un grand nombre de situations de mouvement fréquemment rencontrées dans les séquences d'images. Ces deux algorithmes étant complémentaires, il serait intéressant de mettre au point une stratégie de passage de l'un à l'autre selon la séquence d'images traitée.

Une autre voie d'investigation concerne l'utilisation d'ondelettes 3D pour la construction de la pyramide multirésolution afin de conserver les informations haute fréquence qui sont perdues avec le filtrage passe-bas. Les recherches portent sur la manière d'utiliser cette information haute fréquence (information sur les contours des objets, processus de lignes) dans le but d'améliorer l'analyse du mouvement d'une scène.

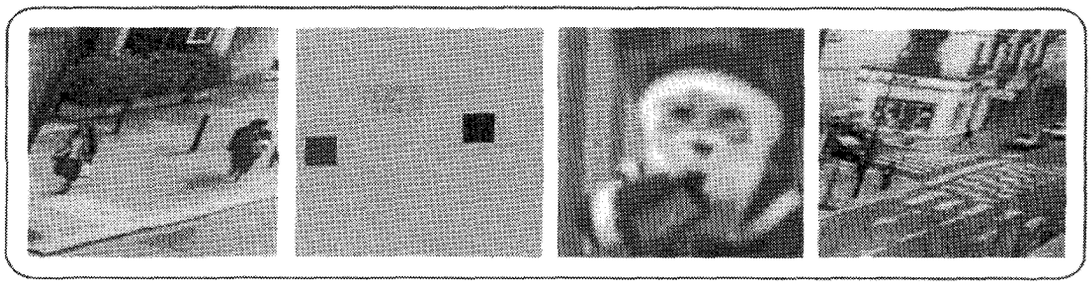

Figure 18. – Une image de chacune des séquences (1,2,3,4) traitées pour l'obtention des résultats du tableau 1.

BIBLIOGRAPHIE

[1] Ayer S., Schroeter P. «Multiple Motion Estimation by Robust Parameter Estimation over Multiple Frames», In Signal Processing VII : Theories and applications, M. Holt, C.Cowan, P. Grant, W. Sandham (Eds.), 1994, pp. 700-703.

[2] Azencott R. «Image Analysis and Markov Random fields», Proc. International Conference on Industrial and Applied Mathematics, SIAM, Philadelphia, 1988, pp. 53-61.

[3] Baaziz N. «Approches d'estimation et de compensation de mouvement multirésolutions pour le codage de séquences d'images», Thèse de doctorat, Université de Rennes I, 1991.

[4] Besag J. «On the Statistical Analysis of Dirty Pictures», In Journal of Royal Statistical Society, Vol.B-48, N.3, 1986, pp.259-302.

[5] Bouman C., Shapiro M. «A Multiscale Random Field Model for Bayesian Image Segmentation», IEEE Trans. on Image Processing, Vol.3, N.2, March 1994, pp.162-177.

[6] Bouthémy P., Lalande P. «Recovery of moving object masks in an image sequence using local spatiotemporal contextual information», Optical Engineering, Vol.32, N.6, June 1993, pp.1205-1212.

[7] Burt P.J., Adelson E.H. «The Laplacian Pyramid as Compact Image Code», IEEE Trans. on Communications, COM-31, N.4, 1984, pp.532-540.

[8] Caplier A. «Modèles markoviens de détection de mouvement dans les séquences d'images : approche spatio-temporelle et mises en œuvre temps réel», Thèse de doctorat de l'INPG, Grenoble, France, Décembre 1995.

[9] Caplier A., Dumontier C., Luthon F., Coulon P.Y. «Algorithme de détection de mouvement par modélisation markovienne. Mise en œuvre sur DSP», A paraître dans *Traitement du Signal*, Vol.13, N.2 1996, 14 pages.

[10] Cross G.R., Jain A.K. «Markov Random Field Texture Models», IEEE Trans. on PAMI, Vol.5, N.1, January 1983, pp.25-39.

[11] Geman S., Geman D. «Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images», In IEEE Trans. on PAMI Vol.6, N.6, November 1984, pp.721-741.

[12] Hsu Y.Z., Nagel H.H., Reckers G. «New Likelihood Test Methods for Change Detection in Image Sequences», Computer Vision, Graphics and Image Processing, CVGIP-26, 1984, pp.73-106.

[13] Lalande P. «Détection du mouvement dans les séquences d'images selon une approche markovienne; application à la robotique sous-marine», Thèse de doctorat, Université de Rennes I, 1990.

[14] Mallat S. «A theory for multiresolution signal decomposition : the wavelet representation», IEEE Trans. Pattern Anal. and Machine Intel., Vol.11, N.7, July 1989, pp.674-693.

[15] Odobez J.M. «Estimation, détection et segmentation du mouvement : une approche robuste et markovienne», Thèse de l'Université de Rennes I, France, Décembre 1994.

[16] Pérez P. «Champs markoviens et analyse multirésolution de l'image : application à l'analyse du mouvement», Thèse de doctorat de l'unversité de Rennes I, France, Juillet 1993.

Manuscrit reçu le 25 juin 1996.

LES AUTEURS

Alice CAPLIER

Alice Caplier est diplômée de l'ENSIEG depuis 1991 et docteur de l'Institut National Polytechnique de Grenoble depuis 1995. Ses travaux de recherche au TIRF concernent l'analyse du mouvement dans les séquences d'images par utilisation d'approches statistiques (champs de Markov) et/ou d'approches multirésolution. Elle travaille aussi sur des mises en œuvre temps réel d'un algorithme de détection de mouvement.

Franck LUTHON

Franck Luthon est maître de conférence à l'ENSERG/INPGrenoble. Ses travaux de recherche au TIRF portent sur l'analyse du mouvement et la compression de séquences d'images.