Suivi d'objets vidéo par propagation d'étiquettes et rétro-projection

Video object tracking by label propagation and backward projection

Guillaume Foret, Pascal Bertolino, Jean-Marc Chassery
Laboratoire des Images et des Signaux, BP 46, 38402 Saint Martin d'Hères, France
Guillaume.Foret, Pascal.Bertolino, Jean-Marc.Chassery@lis.inpg.fr
Tel. : 04.76.82.62.73, Fax. : 04.76.82.63.84

Manuscrit reçu le 5 mars 2004

Résumé et mots clés

**Cet article présente nos travaux sur le suivi d'objets dans un plan séquence.
Un état de l'art sur les techniques de segmentation spatio-temporelle nous permet d'introduire notre propre méthode de suivi temporel d'objets. Elle est constituée de trois phases distinctes : une prédiction d'étiquettes par projection de partition, une segmentation locale associée à une propagation d'étiquettes, et une classification par rétro-projection. L'association de ces trois étapes cumule les avantages de chaque approche pour un suivi rigoureux d'objets et réduit le temps de traitement de chaque image.
La qualité visuelle des résultats obtenus par cette méthode est illustrée en fin d'article. Pour cela nous avons considéré le suivi d'objets ayant des caractéristiques différentes au niveau de leur composition et de leur déplacement.**

Segmentation locale, pyramide irrégulière, propagation d'étiquettes, projection de partition, suivi temporel, objets vidéo.

Abstract and key words

This paper presents an approach dedicated to the tracking of one or several semantic objects in a video shot.
A state of the art on spatio-temporal segmentation techniques allows us to introduce our own approach. It combines three different steps: label prediction based on partition projection, local segmentation associated with a label propagation, and classification by backward projection.
Experimental results highlight the visual quality obtained with this method. Different kinds of objects can be accurately tracked in different kinds of video sequences.

Local segmentation, irregular pyramid, label propagation, partition projection, tracking, video objects.

1. Introduction

Dans le domaine cinématographique, un plan séquence est par définition une suite d'images enregistrées au cours d'une même prise de vues (figure 1). Il existe donc une continuité spatio-temporelle dans le contenu des images d'un plan séquence. De manière intuitive, ce contenu est souvent perçu comme étant composé d'un ou plusieurs objet(s) évoluant sur un arrière-plan donné. La distinction entre la notion d'objet et d'arrière-plan est très subjective, l'arrière-plan pouvant être lui-même composé d'objets.

La segmentation d'un plan séquence suivant son contenu en objets est une opération délicate et difficile à automatiser. Elle nécessite tout d'abord de définir quels sont les objets d'intérêt par rapport à l'arrière-plan. Nous allons considérer dans cet article que cette information est déjà disponible pour la première image $I(t=0)$ (figure 1.a). Plus précisément, ceci pré-suppose que nous possédons une partition $P(0)$ (figure 2), dans laquelle une étiquette est attribuée à chaque pixel de $I(0)$ en fonction de l'objet auquel il appartient.

De manière à simplifier la discussion, nous nous limiterons au suivi d'un seul objet. L'extension de la méthode proposée au suivi de plusieurs objets est immédiate (*cf.* figure 15). De plus aucune contrainte particulière sur la forme, la texture, ou le mouvement des objets considérés n'est imposée. Ceci offre un système générique utilisable dans une gamme variée d'applications.

L'association de l'image $I(0)$ et de la partition $P(0)$ permet de savoir quels sont les éléments de l'image qui constituent l'objet (figure 3). L'objectif est alors de segmenter automatiquement l'objet dans les images suivantes.

Les changements temporels entre deux images successives étant de faible amplitude, une solution largement adoptée est de construire la partition $P(t+1)$ en fonction du résultat précédent $P(t)$. Cette opération est effectuée en utilisant à la fois des informations spatiales (couleurs) et temporelles (mouvement) mesurées sur $I(t)$ et $I(t+1)$; on parle alors de segmentation spatio-temporelle. Les deux principales difficultés rencontrées lors de la segmentation de l'objet dans chaque nouvelle image (à partir de $t=1$) sont :

Figure 2. Exemple d'une partition en objets P(0) associée à l'image I(0) : les deux étiquettes (couleurs) différencient les pixels de l'objet et ceux de l'arrière-plan.

Figure 3. Représentation de l'objet à t = 0.

- La segmentation des zones faiblement contrastées au niveau du contour de l'objet,
- La gestion automatique des nouveaux éléments apparaissant dans l'image.

La première de ces difficultés peut être résolue en appliquant volontairement une sur-segmentation spatiale. Pour répondre à la seconde difficulté, nous faisons l'hypothèse selon laquelle les caractéristiques colorimétriques des éléments découverts sont en accord avec celles de l'objet auquel ils appartiennent.

À la suite d'un état de l'art sur les techniques de segmentation spatio-temporelle, orientées vers le suivi temporel d'objets vidéo, nous introduisons notre propre approche. Les parties suivantes présentent en détails les formalismes retenus en justifiant leur choix. La dernière partie décrit et commente les résultats de suivi d'objets obtenus.

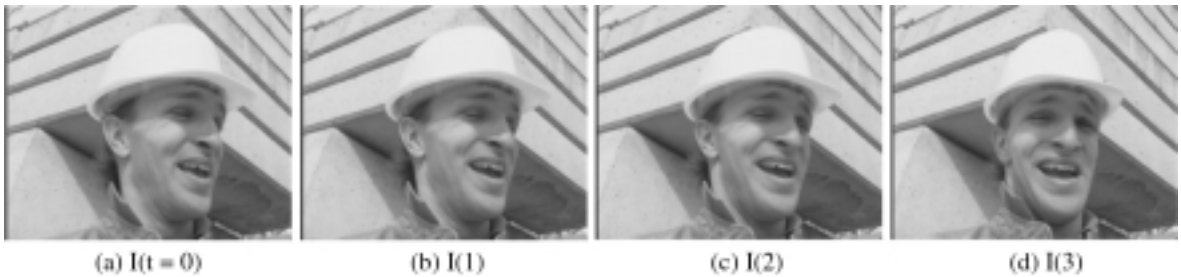

Figure 1. Images originales successives issues d'un même plan séquence (Foreman).

2. Introduction de la méthode

Grand nombre de méthodes ont été proposées pour déduire la partition courante d'une image à partir d'une autre partition. Le cas le plus couramment considéré est celui de deux images successives $I(t)$ et $I(t+1)$, supposant $P(t)$ connue. Deux catégories de méthodes peuvent être distinguées :

1. Les méthodes caractérisées par une **projection** en avant de la partition : $P(t)$ est compensée en mouvement, puis appliquée directement sur l'image $I(t+1)$. Un **ajustement** de cette partition suivant le contenu de $I(t+1)$ permet d'obtenir $P(t+1)$ [Gu 98b, Park 00, Marq 97]. Cet ajustement est réalisé **localement** par un algorithme de segmentation spatiale. L'inconvénient de ces méthodes est de privilégier, lors de l'ajustement, le contour le plus contrasté dans les zones remises en cause. Même si ce contour correspond très souvent au contour réel de l'objet, il peut s'avérer néfaste de ne pas considérer les autres possibilités de segmentation dans ces zones. Une légère délocalisation du contour dans $I(t+1)$ peut entraîner des dégénérescences importantes dans les partitions suivantes.
2. Les méthodes, plus récentes, qui proposent d'appliquer une **segmentation** spatiale sur l'ensemble de $I(t+1)$: la segmentation de l'objet est effectuée en étiquetant les régions segmentées suivant deux classes (objet et arrière-plan) [Alat 98, Marq 98, Gu 98a, Gati 99, Pate 00]. Cette **classification** est réalisée en fonction de la représentation précédente de l'objet. L'inconvénient ici est le coût de calcul nécessaire au traitement de chaque image. En effet la segmentation de la totalité de l'image, ainsi que la classification de chaque région obtenue alourdissent le traitement.

Afin de pallier les inconvénients cités précédemment nous avons étudié l'association de ces deux types d'approches. Nous avons abouti à une technique originale utilisant des outils éprouvés, conçue en trois étapes séquentielles :

- Une projection en avant de $P(t)$,
- Une segmentation spatiale appliquée localement dans $I(t+1)$,
- Une classification des segments locaux obtenus.

Le schéma de la figure 4 présente l'enchaînement de ces trois étapes (modules) permettant de construire la partition $P(t+1)$ à partir de $P(t)$ et des deux images originales $I(t)$ et $I(t+1)$. Le fonctionnement de chacun de ces modules est détaillé dans les parties suivantes en insistant sur leur intérêt respectif.

3. Projection de partition

3.1. État de l'art

L'objectif est de projeter la partition $P(t)$ sur l'image $I(t+1)$ en fonction des similarités mesurées entre $I(t)$ et $I(t+1)$. Cette démarche permet de réduire le temps nécessaire au traitement de $I(t+1)$ et implique une cohérence entre $P(t)$ et $P(t+1)$ favorable à la stabilité du suivi temporel d'un objet vidéo.

Sa réalisation dépend de la manière dont est modélisée la partition $P(t)$. Lorsque cette dernière est représentée par une approximation polygonale de ses contours [Wu 95, Bonn 98, Mahb 02], la projection est appliquée sur les sommets des polygones en fonction du mouvement estimé sur les régions polygonales associées. On déplace ainsi les contours de la partition. Plus couramment $P(t)$ représente la segmentation de $I(t)$ en attribuant à chaque pixel l'étiquette d'une région ou d'un objet. Dans ce cas, la projection en avant de $P(t)$ consiste à prédire la répartition des étiquettes dans $I(t+1)$. Elle peut être effectuée en déplaçant les régions [Wan 98] ou l'objet [Gu 98b, Park 00, Jeha 01] suivant un vecteur mouvement estimé. Une autre solution est de compenser en mouvement $P(t)$ en considérant non pas des régions mais des blocs de pixels [Pard 94b, Marq 97]. L'estimation de mouvement est alors réalisée par mise en correspondance de blocs (block-matching).

Utilisant par la suite un algorithme de segmentation orienté régions (*cf.* section 4), nous avons choisi une représentation par

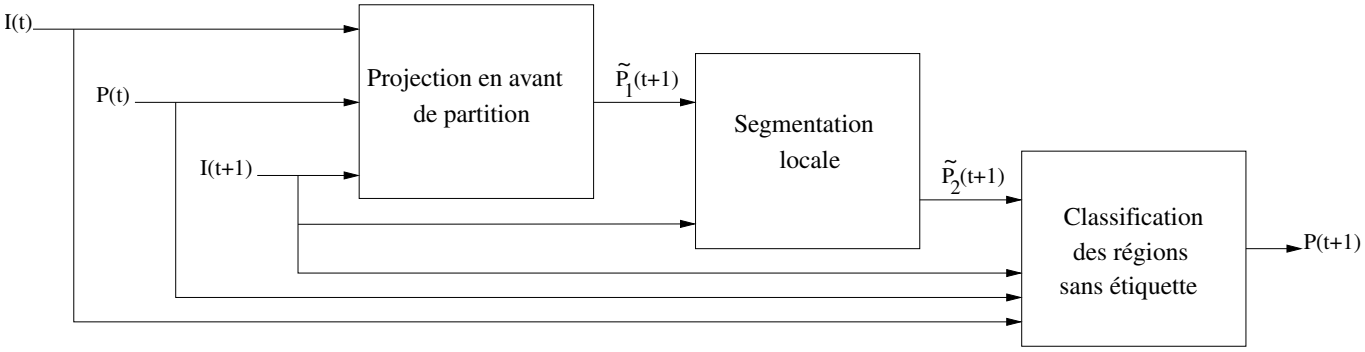

Figure 4. Schéma blocs de la méthode proposée ($\widetilde{P}_1(t+1)$ et $\widetilde{P}_2(t+1)$ correspondent à des partitions intermédiaires).

étiquettes de la partition. La compensation en mouvement des régions ou des objets après avoir estimé leur déplacement entre $I(t)$ et $I(t+1)$ est une approche intuitive. Son inconvénient majeur est le temps de calcul nécessaire à l'estimation de mouvement de chaque région.
Il faut noter que l'objectif n'est pas de construire une partition définitive de $I(t+1)$, mais une prédiction de cette partition. L'utilisation de l'algorithme de block-matching permet d'optimiser ce traitement, tout en conservant une qualité très satisfaisante [Marq 98b].
Nous abordons dans les paragraphes suivants la projection de partition à l'aide de l'algorithme du block-matching. Très largement utilisé en codage vidéo, cet algorithme effectue l'estimation de mouvement entre deux images par un découpage en blocs réguliers de l'une d'elles.

3.2. Mise en correspondance par block-matching

Soient I_1 et I_2 deux images d'un même plan séquence, le vecteur de mouvement d'un bloc de l'image I_1 est obtenu en recherchant dans I_2 le bloc le plus similaire dans une zone limitée de l'image (fenêtre de recherche). Le critère de mise en correspondance de deux blocs est généralement la somme en valeur absolue des différences en niveaux de gris ou en couleurs (*SAD : Sum of Absolute Difference*). La SAD de deux blocs X et Y ($X \in I_1$, $Y \in I_2$) de $N \times N$ pixels est définie par :

$$SAD(X,Y) = \sum_{i=1}^{N} \sum_{j=1}^{N} |X(i,j) - Y(i,j)| \qquad (1)$$

Pour un bloc source X donné, le bloc Y le plus similaire est celui qui minimise la SAD. Il est également possible de retenir comme critère la somme des différences au carré (*SSD : Sum of Squared Difference*).
Dans la littérature scientifique, de nombreux algorithmes de block-matching existent. Un état de l'art récent et synthétique peut être trouvé dans [Chen 01]. La méthode de référence (*Full Search Algorithm (FSA)*) consiste à calculer de manière exhaustive l'erreur SAD pour tous les déplacements possibles dans un voisinage donné. Elle offre la meilleure estimation de mouvement que nous pouvons obtenir avec un block-matching, mais elle est très coûteuse en temps de calcul.
Il est apparu préférable d'utiliser la méthode intitulée *Block Sum Pyramid Algorithm (BSPA)* [Lee97] [Lin98], qui limite le temps de calcul grâce à un algorithme d'éliminations successives (*Successive Elimination Algorithm (SEA)*), et dont l'estimation est d'aussi bonne qualité que la méthode *FSA*. L'algorithme *SEA* fait appel à une représentation pyramidale de chaque bloc de l'image pour éliminer rapidement les blocs non similaires.
Nous rappelons qu'un algorithme de block-matching est caractérisé par deux paramètres :
- La taille des blocs manipulés ($taille\_bloc$). Ce paramètre dépend de la résolution de l'image et par conséquent du format de la vidéo. Pour une séquence CIF 352×288 pixels (respectivement QCIF 176×144 pixels), nous utilisons des blocs de 16×16 pixels (respectivement 8×8 pixels).
- Le vecteur de déplacement maximum autorisé en pixels (V_{maxBM}). Ce paramètre fixe la taille de la fenêtre de recherche. En général, $V_{maxBM} = (8,8)$ ou $(16,16)$.
Lors de l'estimation de mouvement par block-matching, c'est le niveau de gris des pixels qui est utilisé dans le calcul de la SAD.

3.3. Fiabilité des mises en correspondance

Deux blocs sont mis en correspondance suivant la répartition des niveaux de gris de leurs pixels. L'hypothèse selon laquelle tous les pixels d'un bloc effectuent le même mouvement de translation constitue à la fois la force et le point faible du block-matching. Cette hypothèse peut en effet s'avérer restrictive, notamment lorsqu'une rotation intervient dans l'image. Elle peut également ne pas être vérifiée lorsqu'un bloc contient des morceaux de plusieurs régions, et que ces dernières ont des mouvements différents. Cependant, dans les vidéos traitées, ces limites sont peu contraignantes compte tenu de la faible amplitude du mouvement.
La mise en correspondance de deux blocs est réalisée en minimisant la valeur de la *SAD*. La valeur *SAD* minimum obtenue peut être utilisée pour évaluer la fiabilité d'une mise en correspondance. Elle peut, par conséquent, servir à éviter une erreur d'estimation de mouvement.
Lorsque cette valeur est faible, le vecteur de déplacement déduit est donc fiable (en particulier dans le cas de blocs hétérogènes). En revanche, lorsque la valeur minimum de la *SAD* est supérieure à un seuil de fiabilité (T_{SAD}), il est préférable de ne pas tenir compte de cette mise en correspondance. Le bloc en question ne pourra pas être utilisé pour l'étape de prédiction.

3.4. Application à la prédiction de partition

Afin de limiter les erreurs de prédiction, le seuil de fiabilité doit être assez strict (ex : $T_{SAD} = 5$ pour des blocs de 8×8 pixels). Comme le montre le paragraphe suivant, ce seuil peut varier en fonction du contenu et du niveau de bruit dans la scène traitée.
L'estimation de mouvement entre $I(t)$ et $I(t+1)$ peut être effectuée dans les deux sens temporels $t \rightarrow t+1$ (en mode avant) ou $t+1 \rightarrow t$ (en mode inverse). En mode avant, la division en blocs réguliers est appliquée sur l'image $I(t)$. Un vecteur de mouvement est estimé pour chaque bloc. Le même découpage par blocs est appliqué sur la partition $P(t)$. Chaque bloc source de $P(t)$ est alors projeté suivant son vecteur mouvement, afin de prédire la partition de $I(t+1)$ (figure 5.a). Le résultat obtenu contient des zones non recouvertes, dues soit à la superposition des blocs lors de leur projection, soit au seuil de fiabilité. Aucune prédiction n'est retenue lorsqu'un pixel est associé à plusieurs objets.
À cette prédiction de partition, il est préférable d'utiliser celle obtenue avec l'algorithme de block-matching en mode inverse

Figure 5. Illustration des résultats obtenus lors de la projection par block-matching de la partition P(t = 0) (fig. 2) sur l'image I(1) (fig. 1. b)

(division en blocs réguliers appliquée sur $I(t+1)$), car les zones non recouvertes sont ainsi limitées à l'ensemble des blocs sans correspondant. Une prédiction de la partition de $I(t+1)$ est alors obtenue en remplaçant chaque bloc source par le bloc correspondant dans $P(t)$, s'il existe (figure 5.b).

3.5. Analyse et synthèse de la projection

L'objectif principal de cette étape de projection est de simplifier le traitement des étapes suivantes sans introduire d'erreur. C'est pourquoi par défaut $T_{SAD} = 5$ qui est une valeur relativement stricte. Pour certaines séquences, ce seuil peut être augmenté. Par exemple, dans la séquence Coastguard, la présence d'eau dans la scène rend difficile la mise en correspondance entre blocs. Cette difficulté est accentuée en considérant deux images éloignées de la séquence. La figure 6 illustre qu'un seuil plus élevé permet une prédiction plus riche dans ce cas. Cette prédiction facilitera la segmentation dans la nouvelle image. En contrepartie les risques d'erreurs de prédiction sont plus élevés. Dans la figure 5.b, les blocs représentés en noir sont les blocs qui n'ont pu être prédits ; leurs pixels ne possèdent pas d'étiquette. De manière générale, ces zones non prédites peuvent résulter de l'apparition d'un nouvel objet, du découvrement de l'arrière-plan par les objets suivis, ou bien encore de changements importants du contenu de l'image (déformation de l'objet). Citons comme exemple l'oeil gauche du personnage qui se découvre progressivement. Ces zones nécessitent une re-segmentation dans la nouvelle image.

En outre la prédiction des étiquettes dans $I(t+1)$ fournit une approximation du contour de l'objet suivi. Quelques irrégularités sont observables compte tenu de la juxtaposition indépendante des blocs. Les pixels proches de ce contour doivent être resegmentés pour assurer la qualité de segmentation. C'est pourquoi les étiquettes prédites au niveau du contour sur une largeur de 8 à 10 pixels sont supprimées (figure 7).

Figure 7. Prédiction finale des étiquettes dans I(1).

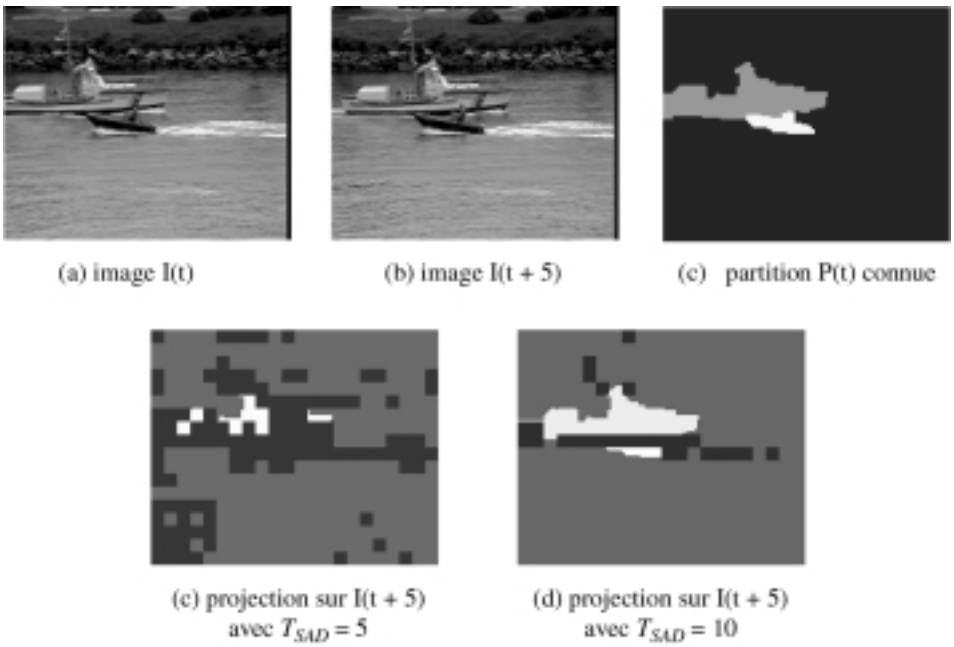

Figure 6. Projection, à partir des images d'origine I(t) et I(t + 5), de la partition P(t) sur l'image I(t + 5) pour deux valeurs de T_{SAD}.

3.6. Conclusion

Nous venons de présenter une méthode adaptée à la projection temporelle de partition. Elle s'appuie sur l'algorithme relativement simple du block-matching. Cette méthode se révèle être efficace concernant les deux objectifs suivants :
- Prédire une étiquette pour un maximum de pixels dans $I(t+1)$, avec un taux de confiance élevé
- Détecter les zones temporellement instables entre $I(t)$ et $I(t+1)$.

La robustesse de cette méthode de projection par rapport aux mouvements de plus forte amplitude sera mise en évidence par des résultats de suivi présentés à la section 6. Nous présentons dans la partie suivante la manière dont sont traitées les zones sans prédiction, l'objectif étant de segmenter en totalité l'objet d'intérêt. Notons que les zones sans prédiction seront appelées par la suite zones d'incertitude [Pard 94].

4. Segmentation spatiale

4.1. État de l'art

Afin d'extraire avec précision le contour de l'objet suivi, une segmentation spatiale est nécessaire dans chaque nouvelle image. Suivant l'approche adoptée, cette segmentation est soit initialisée par la projection de la partition précédente (segmentation locale) soit appliquée de manière indépendante sur l'ensemble de l'image (segmentation globale).

Concernant la segmentation locale, un grand nombre de travaux ont retenu l'algorithme de la ligne de partage des eaux [Vinc 91]. Cet algorithme est fondé sur un principe d'agrégation de pixels autour d'attracteurs. Il est en effet bien adapté au problème d'ajustement de la partition projetée : les zones d'incertitude sont segmentées en définissant des attracteurs dans les zones étiquetées de l'image [Marq 97, Wang 98, Gu 98b, Park 00]. Soulignons ici que la théorie des contours actifs apporte également une solution pour ajuster le contour de l'objet projeté dans $I(t+1)$ [Jeha 01]. Cependant cette approche ne permet pas de gérer les zones temporellement instables autres que celles liées aux déformations du contour lui-même.

Lorsque la segmentation spatiale est appliquée sur l'ensemble de l'image $I(t+1)$, les algorithmes utilisés sont plus variés : ligne de partage des eaux [Marq 98, Gati 01a], approche stochastique [Pate 00], segmentation par croissance de régions [Alat 98, Gu 98a]. L'objectif est une sur-segmentation de l'image traitée, de manière à extraire l'ensemble des contours. Les régions segmentées sont attribuées ou non à l'objet suivant le résultat d'une projection temporelle. Le contour de l'objet segmenté dépend alors de la classification de ces régions.

Au cours de la section précédente, nous avons montré qu'il est possible de prédire de manière fiable et rapide une étiquette pour un grand nombre de pixels de $I(t+1)$. La segmentation de la totalité de cette image est donc inutile et serait une perte de temps. De plus le résultat de la prédiction fournit des informations pertinentes pour guider la segmentation spatiale dans $I(t+1)$. Mais, contrairement aux méthodes existantes fondées sur un ajustement, nous préférerons réaliser une segmentation en régions homogènes des zones d'incertitude. Le contour de l'objet est ainsi localisé en considérant l'ensemble des hétérogénéités présentes dans ces zones. Pour cela nous effectuons une segmentation spatiale fondée sur le principe de la **pyramide irrégulière** [Mont 91].

4.2. Segmentation par pyramide irrégulière

Issue du domaine de la théorie des graphes, la pyramide irrégulière est une technique de croissance de régions en parallèle. L'utilisation de cette structure pour la segmentation en régions homogènes d'une **image** se réalise de manière hiérarchique. Considérant l'ensemble des pixels de l'image comme étant un ensemble de régions, cette technique initialise un graphe d'adjacence qui modélise la 4 ou 8 connexité avec un sommet par pixel. Ce graphe symbolise le premier niveau de la pyramide ; les niveaux suivants sont obtenus en simplifiant localement le graphe d'adjacence en respectant certaines règles, ce qui provoque le regroupement progressif de régions.

Le regroupement de deux régions **adjacentes** est autorisé tant que la distance en couleur calculée entre ces deux régions est inférieure à un seuil fixé T_{seg} (seuil global de similarité). Dans notre cas, cette distance est définie sur les composantes de luminance (y) et de chrominances (u, v) de la manière suivante :

$$d(R_1,R_2) = \sqrt{(y_1-y_2)^2 + \gamma^2.[(u_1-u_2)^2 + (v_1-v_2)^2]} \quad (2)$$

où (y_i, u_i, v_i) représentent les composantes couleurs moyennes YUV de la région R_i. γ est un coefficient normalisateur utilisé pour compenser la différence d'échelle entre la composante de luminance et les deux chrominances. Il est calculé automatiquement sur une image en fonction de la largeur des histogrammes respectifs :

$$\gamma = min\left(\frac{y_{max}-y_{min}}{u_{max}-u_{min}}, \frac{y_{max}-y_{min}}{v_{max}-v_{min}}\right) \quad (3)$$

Une fois construite, la pyramide est un empilement de partitions dont la résolution décroît de la base vers le sommet [Bert 95]. Chacune de ces partitions représente l'image par un ensemble de régions, homogènes en couleur (figure 8). La forme des régions extraites n'est contrainte par aucun critère géométrique, ce qui constitue une particularité de cette approche pyramidale. Notons la possibilité d'utiliser un paramètre supplémentaire $T_{minRegion}$ fixant la taille minimale d'une région dans la partition du dernier niveau de la pyramide.

Par la suite, seul le dernier niveau de la pyramide est considéré comme résultat de la segmentation.

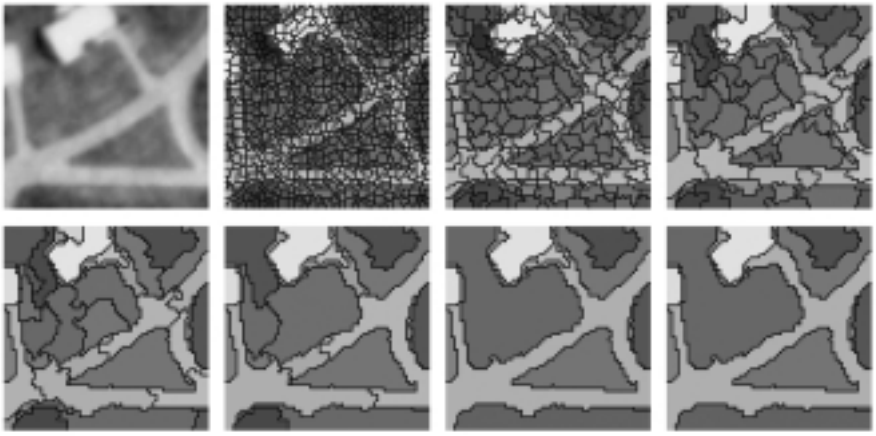

Figure 8. Résultat de segmentation par pyramide irrégulière (extrait de [Bert 95]).

4.3. Influence du seuil global de similarité

Le seuil global de similarité T_{seg} ne varie pas au cours de la construction de la pyramide. Il est fixé par l'utilisateur avant d'exécuter une segmentation. Il correspond à une limite de sécurité au delà de laquelle deux sommets adjacents sont considérés non similaires. Un seuil élevé favorise les fusions entre régions peu similaires. Une étude de l'influence de ce seuil est disponible dans [Bert 95].
Suivant la valeur de T_{seg} choisie, la segmentation par pyramide irrégulière fournit différentes «résolutions» de segmentation (figure 9). L'utilisation d'un seuil trop élevé peut entraîner des imperfections telle que la perte de certains contours (*cf.* figure 9.d). Pour une segmentation précise, un seuil relativement strict est recommandé ($T_{seg} < 10$ pour des séquences test MPEG: Foreman, Mother&Daugther...). La sur-segmentation de l'image peut être limitée grâce au paramètre $T_{minRegion}$.

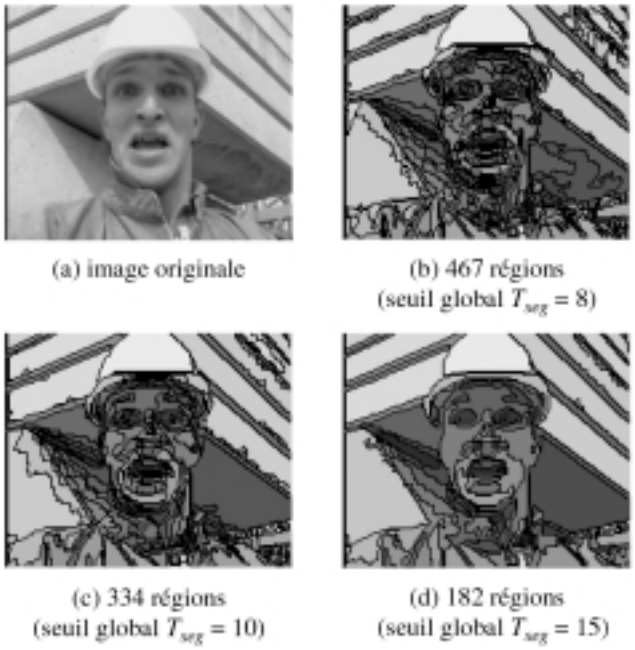

Figure 9. Résultats de segmentation (sommets de la pyramide) obtenus pour différentes valeurs du seuil global de similarité en couleur T_{seg} (avec $T_{minRegion} = 7$ pixels).

4.4. Segmentation locale et propagation d'étiquettes

La modélisation par graphe fournit un cadre algorithmique bien formalisé et une grande souplesse d'utilisation. Ainsi l'algorithme de la pyramide irrégulière peut être appliqué sur des ensembles de pixels de formes arbitraires dans une image, tout en conservant les relations d'adjacence entre les résultats locaux de segmentation et le reste de l'image.
De plus, de par sa structure, la pyramide permet d'introduire facilement pour chaque sommet des informations supplémentaires telles que des étiquettes. Il est donc possible de réaliser une propagation d'étiquettes entre sommets au cours des fusions entre régions adjacentes.
Ces deux caractéristiques font de la pyramide irrégulière un outil adapté à la segmentation locale dans une image à partir d'une partition pré-existante, telle que celle obtenue après une projection de partition (figure 7). L'algorithme de segmentation est alors appliqué sur les pixels sans étiquette et les pixels de leur voisinage proche, déjà étiquetés, afin d'effectuer une propagation d'étiquettes des pixels étiquetés vers les pixels non-étiquetés, lors des fusions [Fore 03b]. Ainsi une majorité des régions segmentées sont rattachées au reste de la partition. Certaines régions peuvent également émerger sans étiquette. Nous ne détaillerons pas plus ici le fonctionnement de cette segmentation locale par pyramide irrégulière. Une illustration de son utilisation est fournie au paragraphe suivant.

4.5. Application au suivi temporel d'objets

La deuxième étape de notre méthode de suivi temporel d'objets vidéo consiste donc à segmenter localement $I(t+1)$ (figure

Figure 10. Exemple de résultat de la propagation d'étiquettes dans la partition présentée à la figure 7. L'algorithme de la pyramide a été appliqué avec les paramètres : $T_{seg} = 7$ et $T_{minRegion} = 10$ pixels. Les régions segmentées sans étiquette sont représentées en noir.

1.b) de manière à traiter les zones d'incertitude résultantes de la projection (figure 7).
Pour ce faire, une segmentation locale par pyramide irrégulière est réalisée sur les zones d'incertitude, en considérant aussi les pixels voisins déjà étiquetés sur une largeur de 2 à 3 pixels. La propagation des étiquettes permet ainsi de segmenter les zones d'incertitude en utilisant la cohérence (continuité) spatiale des objets (figure 10). Les régions non similaires à l'arrière-plan et à l'objet restent sans étiquette. Par la suite, la classification de ces régions (troisième étape) permettra la localisation complète du contour de l'objet.
Une sur-segmentation ($T_{seg} = 7$) est préférable afin d'éviter tout risque de fusions abusives dans les zones faiblement contrastées.

4.6. Importance de la sur-segmentation

L'inconvénient éventuel d'une sur-segmentation est d'obtenir en grand nombre des régions de petite taille. Ceci augmente le temps de traitement nécessaire à la classification. Cependant la segmentation n'étant pas appliquée sur la totalité de l'image, nous pouvons nous permettre de conserver la plupart de ces régions. Seules les plus petites sont éliminées au cours de la segmentation locale par pyramide irrégulière en fonction du paramètre $T_{minRegion}$.
Le choix de ce paramètre est important. La taille minimale des régions manipulées conditionne la précision de la localisation des contours. Expérimentalement, des valeurs $T_{minRegion}$ supérieures à 10 pixels entraînent localement des imprécisions, dues à une propagation forcée d'étiquettes. Ces imprécisions sont souvent accentuées au cours du suivi temporel et rendent inutilisables les résultats.
Des valeurs relativement faibles pour $T_{minRegion}$ (ex: 5 pixels) sont recommandées. Il faut noter de plus que les erreurs de classification concernant les régions de petites taille peuvent être facilement corrigées par le post-traitement morphologique. Ce dernier lisse localement la partition finale et limite les éventuelles imperfections dues à une mauvaise classification. $T_{minRegion}$ ne doit cependant pas être trop faible sinon la mise en correspondance devient aléatoire car trop sensible au bruit.

4.7. Conclusion

En conclusion cette deuxième étape présente deux utilités :
- l'ajustement du contour de l'objet en segmentant en régions la fine couche de pixels remise en cause au voisinage du contour prédit,
- le traitement des zones temporellement instables en les segmentant également en régions homogènes.
La segmentation est effectuée en prenant en compte les étiquettes prédites afin de rattacher au mieux les morceaux segmentés au reste de la partition. Certaines régions restent cependant sans étiquette. Leur classification à l'intérieur de l'objet ou de l'arrière-plan est présentée dans la partie suivante.

5. Classification par rétro-projection

5.1. État de l'art

L'objectif de cette troisième et dernière étape est d'attribuer une étiquette à chaque région non étiquetée. Ces régions sont classées suivant la partition précédente de manière à favoriser la cohérence temporelle dans la description de l'objet suivi.
Les approches présentées dans [Marq 98b, Alat 98] effectuent la classification de régions segmentées dans $I(t+1)$ en fonction de la projection en avant de $P(t)$. Dans notre approche, le résultat de cette projection est utilisé pour guider la segmentation. Il ne peut donc pas servir à la classification des régions sans étiquette.
Une classification par projection en arrière (rétro-projection) des régions sans étiquette est réalisée : le mouvement de chaque région est estimé et chacune est projetée sur la partition précédente afin de déterminer si elle appartient ou non à l'objet suivi [Gu 98a, Gati 99, Pate 00]. Afin d'être plus robuste aux changements d'aspect de l'objet suivi, plusieurs partitions, correspondant à différents instants de la vidéo, pourraient être utilisées [Gati 01b]. Cette solution particulière n'a pas été retenue. La réinitialisation de la partition par un utilisateur semble préférable et moins complexe lors d'un changement trop important de l'objet.
La précision obtenue par ces méthodes nous a incités à faire appel au principe de la projection en arrière pour classer les régions sans étiquette.
C'est d'ailleurs le choix retenu dans l'approche de [Mahb 02] pour classer les zones découvertes, tandis que les déformations d'objet et les mises en mouvement sont gérées par une analyse plus poussée du mouvement estimé lors de la projection des régions.

5.2. Estimation du mouvement des régions

La projection des régions sans étiquette requiert l'estimation de leur vecteur de mouvement de $I(t+1)$ vers $I(t)$. À l'instar de [Gu 98a], un modèle de mouvement translationnel permet d'estimer le mouvement d'une région. Chaque région sans étiquette, dans $I(t+1)$, est alors mise en correspondance (region-matching) dans $I(t)$.
Contrairement au block-matching utilisé lors de la première étape, l'information couleur (y,u,v) de chaque pixel p est prise en compte pour renforcer l'exactitude de la classification. Soit R une région de $I(t+1)$, son vecteur de mouvement V est obtenu en minimisant la Somme en valeur Absolue, pour tous les pixels $p \in R$, des Différences en couleur (SAD_{color}):

$$SAD_{color}(R) = \sum_{p \in R} [|(y(t+1,p) - y(t,p+V)| + \gamma. \quad (4)$$
$$(|u(t+1,p) - u(t,p+V)| + |v(t+1,p) - v(t,p+V)|)]$$

L'estimation du vecteur V est effectuée en considérant un vecteur de déplacement maximal en pixels $V_{maxRegion}$ qui fixe les dimensions de la fenêtre de recherche (ex: $V_{maxRegion} = (16,16)$).
Soulignons ici que nous nous intéressons principalement à la mise en correspondance, et non pas à la qualité de l'estimation de mouvement obtenue.

5.3. Classification des régions sans étiquette

Lors de leur projection en arrière sur $P(t)$ (figure 11.a) chaque région se voit attribuer l'étiquette qu'elle recouvre majoritairement. À l'issue de ce traitement, une partition totale de $I(t+1)$ est obtenue (figure 11.b).
Un traitement morphologique simple (une fermeture, puis une ouverture sur le masque de l'objet) est appliqué à la partition finale de manière à lisser les contours et améliorer la qualité visuelle (figure 11.c).

5.4. Bilan sur la classification par rétro-projection

Une région sans étiquette correspond dans $I(t+1)$ à une entité ayant ses propres caractéristiques en couleurs.
Dans la plupart des cas, cette entité est présente dans $I(t)$. Le fait qu'elle n'ait pas d'étiquette est dû soit à sa petite taille soit à une légère déformation. Au cours de la rétro-projection, ces entités sont mises en correspondance avec leur propre représentation dans l'image précédente et peuvent ainsi récupérer la bonne étiquette. Ceci facilite la stabilité temporelle dans la localisation du contour de l'objet.
Lorsqu'une entité n'est pas présente dans $I(t)$. Elle est projetée par défaut sur une partie qui lui est similaire en couleur. Nous faisons alors l'hypothèse que cette similarité en couleur est suffisante pour lui attribuer une étiquette. Il est à noter que face aux problèmes dus aux découvrements, la classification par rétro-projection est plus pertinente que la propagation d'étiquettes. En effet, l'étiquetage forcé d'une région au cours de l'étape de segmentation spatiale l'attribue par défaut à la région la plus similaire de son voisinage. La classification par rétro-projection permet de ne pas se limiter uniquement au voisinage proche pour classer une région.
L'utilisation des trois composantes couleur lors de la mise en correspondance des régions renforce l'exactitude de la classification. Faute de cartes de vérité terrain pour les vidéos étudiées, une quantification de l'apport de l'information couleur n'est actuellement pas possible. Néanmoins, la qualité visuelle de la segmentation est bien améliorée comme le montre le test présenté à la figure 12. La classification des régions segmentées dans $I(t+1)$ (fig. 12.d) échoue à plusieurs reprises lorsque la rétro-projection n'utilise que l'information de luminance (fig. 12.e), contrairement au cas où l'information couleur est utilisée (fig. 12.f).

Figure 11. Classification des régions sans étiquette par rétro-projection.

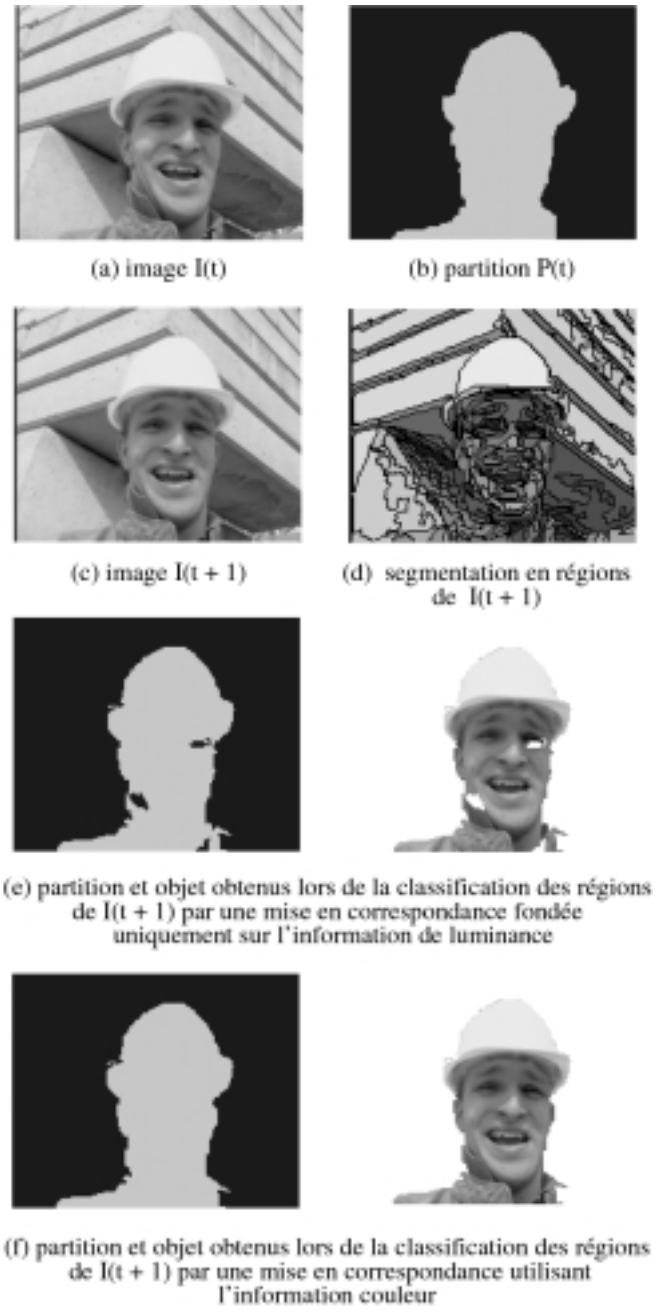

Figure 12. Résultat d'un test illustrant l'apport de l'information couleur par rapport à l'information de luminance seule lors de la mise en correspondance de régions. La classification des régions segmentées dans I(t + 1) (d) échoue à plusieurs reprises lorsque la rétro-projection n'utilise que l'information de luminance (e), contrairement au cas où l'information couleur est utilisée (f).

6. Résultats

Nous présentons dans cette dernière partie des résultats de suivi temporel obtenus avec des séquences test du standard MPEG au format QCIF (176×144) : Foreman, Coastguard, Mother&Daughter et Carphone. Pour chacune de ces séquences, la partition initiale en objets $P(0)$ (ex. : fig. 13.a, fig. 14.a et fig. 15.a) a été obtenue à l'aide d'une interface graphique interactive [Fore 03].

Après avoir mis en évidence la qualité du suivi sur des séquences originales, nous nous intéresserons au comportement de la méthode sur des séquences modifiées obtenues par sous-échantillonnage temporel. Nous illustrerons enfin par un exemple une limite d'utilisation de la méthode.

6.1. Suivi dans les séquences originales

Les résultats portent ici sur trois suivis, effectués sur une cinquantaine d'images successives, avec le même jeu de paramètres :

1. Projection de partition par block-matching :
   - taille des blocs utilisés : $taille\_bloc = 8$ pixels
   - vecteur de déplacement maximum autorisé en pixels : $V_{maxBM} = (8,8)$
   - seuil de fiabilité de mise en correspondance : $T_{SAD} = 5$
2. Segmentation locale par pyramide irrégulière :
   - seuil global de similarité : $T_{seg} = 7$
   - taille minimale autorisée d'une région : $T_{minRegion} = 5$ pixels
3. Classification par projection en arrière :
   - vecteur de déplacement maximum autorisé en pixels pour une région : $V_{maxRegion} = (16,16)$

La figure 13 montre un suivi de très bonne qualité visuelle pour un objet non rigide. Les déformations de l'objet, ainsi que le mouvement de la caméra, entraînent l'apparition de nouveaux éléments au cours du traitement (exemple : la partie gauche du visage). Ces éléments découverts sont ici attribués correctement à l'objet ou à l'arrière-plan. Leurs caractéristiques spatiales (couleur) ont permis leur mise en correspondance avec les parties déjà visibles de l'objet auquel ils appartiennent.

De manière générale la gestion automatique des éléments découverts est délicate et source d'erreur. C'est pourquoi l'interruption du suivi temporel doit être proposée à l'utilisateur, afin de lui permettre de réinitialiser la partition P à un instant particulier. Le manque d'information sémantique sur les zones découvertes conduit inévitablement à ce constat. On peut noter toutefois que le suivi proposé peut être réalisé sans interruption sur plus d'une cinquantaine d'images, avec des variations d'aspect non négligeables de l'objet suivi.

La figure 14 fournit le résultat du suivi d'un objet composé de nombreuses régions homogènes de petite taille. Ce résultat illustre la robustesse de la méthode à suivre des contours faiblement contrastés entre l'objet et l'arrière-plan (*cf.* les contours entre les extrémités du bateau et l'eau).

L'approche présentée dans cet article peut également s'étendre au suivi de plusieurs objets (figure 15). Il suffit de fournir une partition initiale $P(0)$ adéquate. Dans ce cas, la représentation par graphe de la pyramide irrégulière est un atout pour modéli-

ser l'interaction entre les objets qui évoluent dans le plan séquence observé.
Sur un PC Pentium III à 1 GHz, des cadences de traitement de 1 à 3 images par seconde sont obtenues, en fonction de la complexité des objets suivis et de leur nombre.

6.2. Suivi dans des séquences sous-échantillonnées temporellement

Afin de valoriser la robustesse de l'algorithme par rapport aux mouvements de forte amplitude et aux déformations brutales, deux nouvelles séquences test ont été construites. La première, nommée ForemanBis, correspond à un sous-échantillonnage temporel avec un pas de 5 de la séquence Foreman précédente. La seconde, CarphoneBis, est un sous échantillonnage avec un pas de 10 de la séquence Carphone.
Les figures 16 et 17 fournissent les résultats de suivi obtenus sur ces deux séquences (les paramètres sont inchangés, excepté $V_{maxBM} = (16,16)$). Nous pouvons y observer l'efficacité de l'étape de projection de partition (deuxième colonne des figures 16 et 17), qui permet de localiser le contour de l'objet dans chaque nouvelle image et de détecter les zones temporellement instables.
Il faut noter que dans ces séquences sous-chantillonnées, le déplacement de l'objet découvre de manière plus importante l'arrière-plan. De plus la composition de l'objet varie plus fortement. La gestion des éléments découverts est alors la principale difficulté. Pour ces deux séquences, la méthode proposée parvient à traiter correctement ces zones d'incertitude.

6.3. Limite d'utilisation de la méthode

La véritable limite de la méthode se situe au niveau de la gestion automatique des éléments découverts dans l'arrière-plan. Certaines séquences (telle que Paris, figure 18), possèdent un arrière-plan complexe très détaillé. Lors de la segmentation locale, un grand nombre d'éléments de l'arrière-plan sont segmentés sans étiquette. Des difficultés apparaissent au niveau de l'étape de classification. En effet la diversité de ces éléments affaiblit l'hypothèse qu'une entité découverte appartient à l'objet avec lequel elle est mise en correspondance dans l'image précédente.
La figure 18 illustre cette remarque à l'aide de la séquence Paris (au format CIF (352×288)). Dans cette séquence, l'arrière-plan est constitué de nombreux objets de petite taille. Des erreurs de segmentation ont lieu lors du déplacement des personnages. Certaines entités découvertes dans l'arrière-plan sont en effet plus similaires au contenu de l'un des deux personnages qu'à l'arrière-plan lui-même. Citons, comme exemple, les livres découverts à l'arrière de la tête de la femme dans l'image I(30). Ces erreurs entraînent des modifications du contour des objets. Face à ce problème, aucune solution efficace n'a pu être apportée à notre méthode, par manque d'informations pertinentes lors de l'étape de classification où seule la couleur est actuellement utilisée. Suivant l'hypothèse selon laquelle une zone découverte est due au déplacement d'un objet, l'exactitude de la classification d'une entité pourrait être renforcée en tenant compte de son mouvement. Ceci sous-entend de retarder la classification de chaque entité de manière à estimer son déplacement. La segmentation d'une image serait alors dépendante à la fois des images précédentes et suivantes. La stabilité d'un tel traitement reste encore à démontrer.

7. Conclusion

L'originalité de notre approche réside dans la combinaison de trois étapes usuelles en segmentation spatio-temporelle : projection/segmentation spatiale/classification. Cette association est encore peu étudiée. Elle cumule pourtant les avantages de chaque opération pour un suivi rigoureux d'objets vidéo, et entraîne une réduction du temps de traitement de chaque image. Au cours de notre étude, l'algorithme du block-matching s'est avéré être un outil simple et efficace pour réaliser la projection de partition entre deux images.
La pyramide irrégulière, quant à elle, répond parfaitement aux besoins d'une segmentation locale. Elle ajuste les contours prédits des objets, et segmente les zones temporellement instables à l'aide d'une propagation d'étiquettes. La conservation de régions sans étiquette à la fin de la segmentation permet de distinguer certaines entités au voisinage des contours des objets.
Le principe de la classification de régions par projection en arrière a été mis en valeur récemment dans de nombreux travaux. Les résultats que nous avons obtenus permettent de confirmer la pertinence d'une telle classification pour finaliser la segmentation des objets suivis.
L'état actuel de la méthode permet d'envisager des applications qui nécessitent des masques d'objets précis. Certaines améliorations restent toutefois à apporter telle que la détection des dégénérescences de la localisation du contour des objets. De plus une étude approfondie des possibilités pour renforcer la robustesse de la classification des nouvelles entités extraites (petites et grandes), nous paraît d'un intérêt majeur. À titre d'exemple, une meilleure prise en compte de l'information mouvement serait un plus pour cette étape. Soulignons toutefois que les solutions adoptées devront être fondées sur des formalismes simples afin de garantir un comportement stable de l'algorithme.
L'une des applications visées concerne la représentation de l'information contenue dans une vidéo. À l'heure actuelle, l'indexation du contenu des images et la construction de résumés par détection de rupture de plans et assemblage d'images clé permettent de répondre partiellement à ce défi technologique. Mais le contenu d'une image clé est souvent lui même trop riche, et les objets d'intérêt sont souvent « minoritaires » dans une image

(par rapport à l'arrière-plan par exemple). Une perspective à plus long terme de notre travail serait donc d'extraire dans une vidéo des objets clé et de construire un dictionnaire interrogeable du contenu de la vidéo.

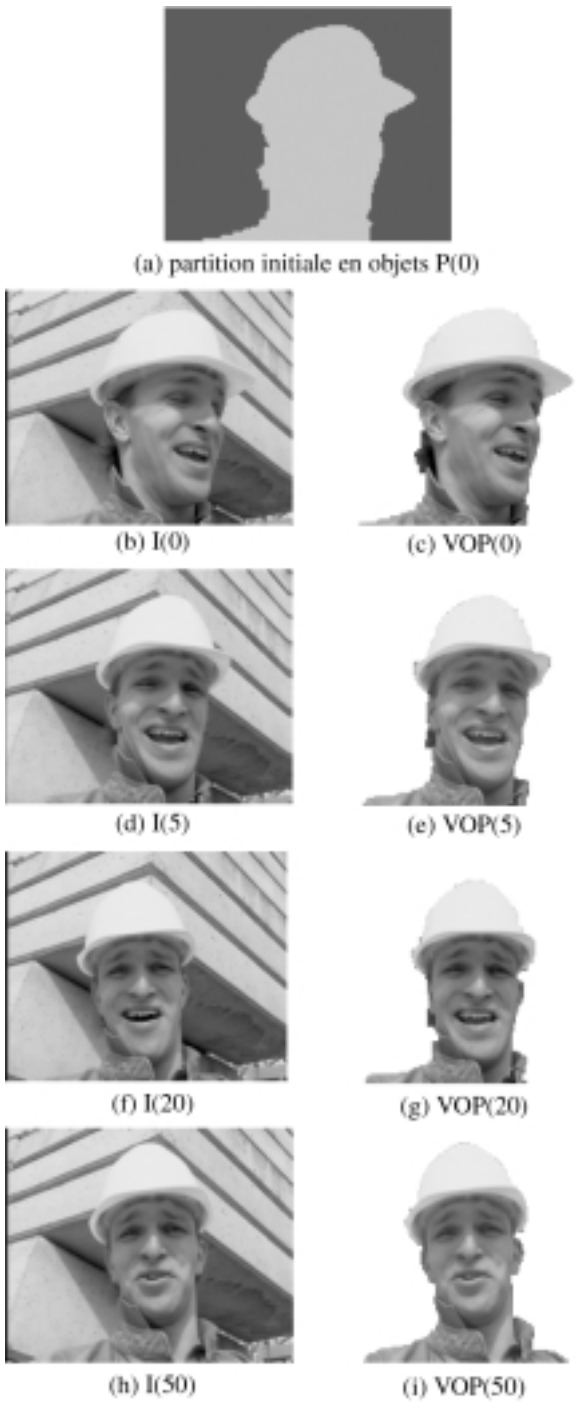

Figure 13. Suivi d'un objet non rigide dans la séquence Foreman (La première colonne présente les images de la séquence à 4 instants différents, la seconde colonne représente l'objet segmenté).

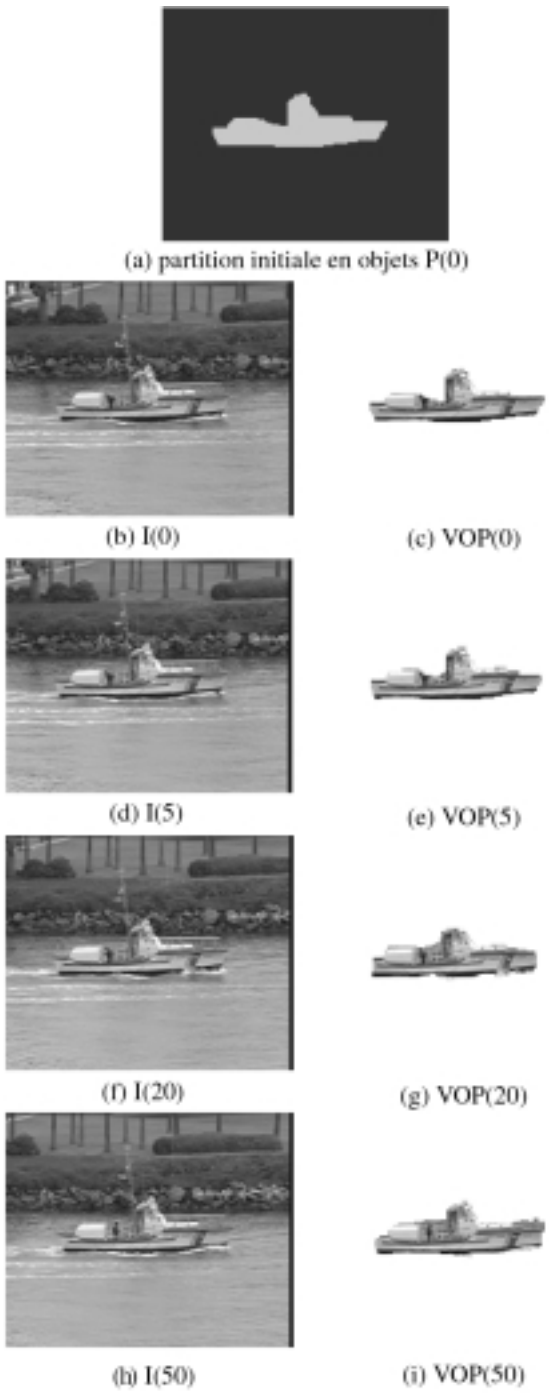

Figure 14. Suivi d'un objet hétérogène dans la séquence Coastguard (La première colonne présente les images de la séquence à 4 instants différents, la seconde colonne représente l'objet segmenté).

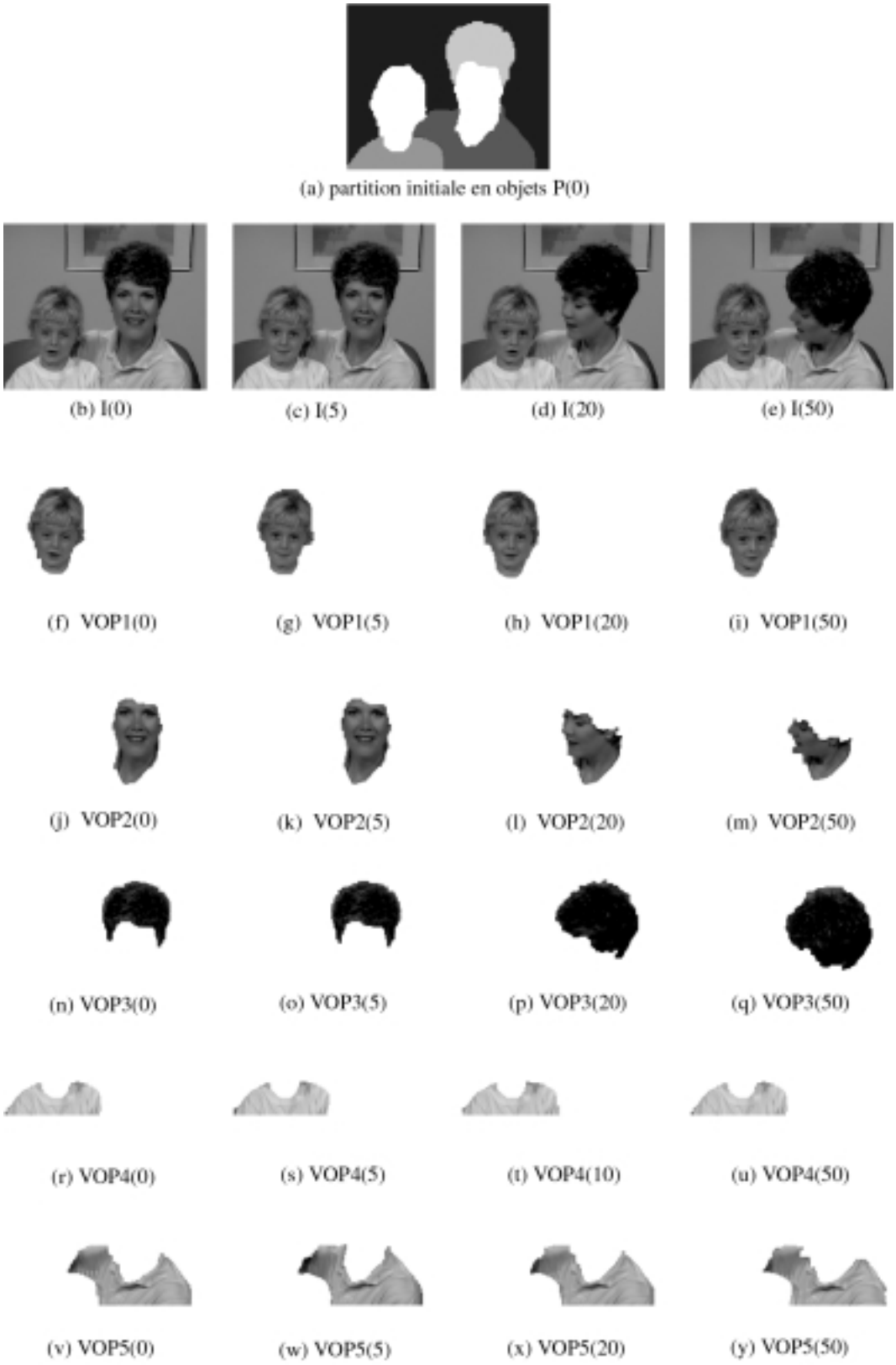

Figure 15. Suivi de plusieurs objets dans la séquence Mother&Daughter (La deuxième ligne représente les images originales à 4 instants différents, les lignes suivantes fournissent les objets segmentés).

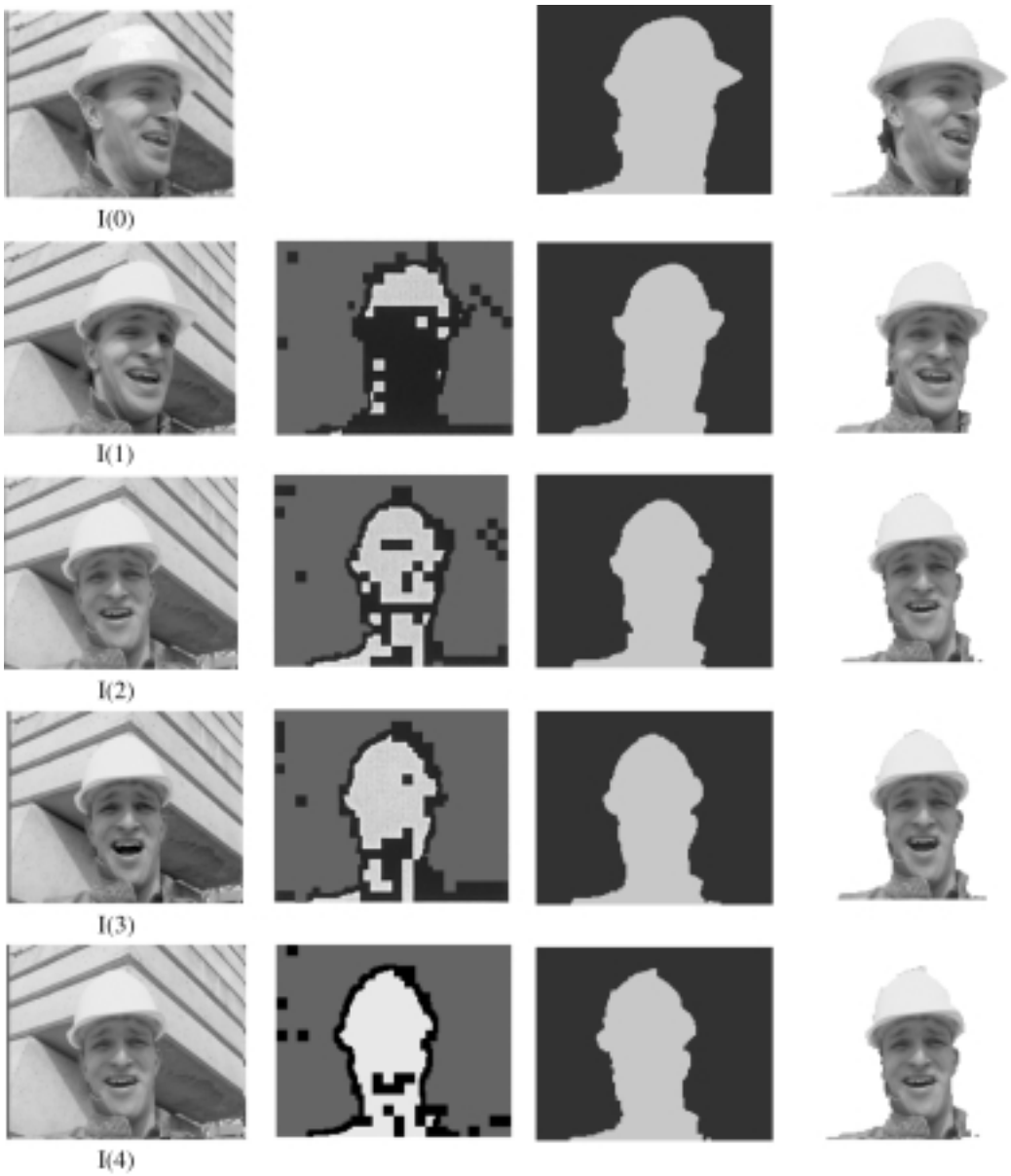

Figure 16. Suivi d'un objet non rigide dans la séquence sous-échantillonnée ForemanBis (La première colonne représente les images successives de la séquence, la seconde fournit le résultat de la projection de partition, la troisième présente la partition en objets obtenue, la dernière donne l'objet vidéo segmenté).

Figure 17. Suivi d'un objet non rigide dans la séquence fortement sous-échantillonnée CarphoneBis (La première colonne représente les images successives de la séquence, la seconde fournit le résultat de la projection de partition, la troisième présente la partition en objets obtenue, la dernière donne l'objet vidéo segmenté)

Figure 18. Segmentation d'objets vidéo évoluant dans la séquence hétérogène Paris (La première colonne représente les images originales à 3 instants différents, les deux autres colonnes donnent les objets vidéo segmentés).

Références

[Alat98] A. ALATAN, L. ONURAL, M. WOLLBORN, R. MECH, E. TUNCEL, T. SIKORA, "Image Sequence Analysis for Emerging Interactive Multimedia Services - The European COST 211 Framework", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 8, N°7, pp. 802-813, November 1998.

[Bert95] P. BERTOLINO, *Contribution des pyramides irrégulières en segmentation d'images multirésolution*, PhD thesis, Thèse de Doctorat de L'Institut National Polytechnique de Grenoble, Novembre 1995.

[Bonn98] L. BONNAUD, *Schémas de suivi d'objets vidéo dans une séquence animée : application à l'interpolation d'images intermédiaires.* PhD thesis, Thèse de Doctorat de l'Université de Renne 1, Octobre 1998.

[Chen01] Y.-S. CHEN, Y.-P. HUNG, C.-S. FUH, "Fast Block Matching Algorithm Based on the Winner-Update Strategy", *IEEE Transactions on Image Processing*, Vol. 10, N°8, pp. 1212-1222, August 2001.

[Fore03] G. FORET, *Segmentation spatio-temporelle d'objets vidéo en vue de leur caractérisation,* PhD thesis, Thèse de Doctorat de L'Institut National Polytechnique de Grenoble, Octobre 2003.

[Gati01a] D. GATICA-PEREZ, C. GU, M. SUN, "Semantic Video Object Extraction Using Four-band Watershed and Partition Lattice Operators", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 11, N°5, pp. 603-618, May 2001.

[Gati01b] D. GATICA-PEREZ, M. SUN, C. GU, "Multiview Extensive Partition Operators for Semantic Video Object Extraction", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 11, N°7, pp. 788-801, July 2001.

[Gati99] D. GATICA-PEREZ, M. SUN, C. GU, "Semantic Video Object Extraction based on Backward Tracking of Multivalued Watershed", In: *Proc. of the IEEE International Conference on Image Processing, ICIP'99*, p., Kobe, Japan, 1999.

[Gu98a] C. GU, M. LEE, "Semantic Video Object Tracking using region-based Classification", In: *Proc. of the IEEE International Conference on Image Processing*, ICIP'98, pp. 643-647, Chicago, USA, 1998.

[Gu98b] C. GU, M. LEE, "Semiautomatic Segmentation and Tracking of Semantic Video Objects'", *IEEE Transactions on Circuits and*

Systems for Video Technology, Vol. 8, N°5, pp. 572-584, September 1998.

[Jeha 01] S. JEHAN-BESSON, M. BARLAUD, G. AUBERT, "Region-based active contours for video object segmentation with camera compensation", In : *Proc. of the IEEE International Conference on Image Processing, ICIP'01*, Thessaloniki, Greece, 2001.

[Lee 97] C. LEE, L. CHEN, "A Fast Motion Estimation Algorithm Based on the Block Sum Pyramid", *IEEE Transactions on Image Processing*, Vol. 6, N°11, pp. 1587-1591, November 1997.

[Lin 98] C. LIN, Y. CHANG, Y. CHEN, "Hierarchical Motion Estimation Algorithm based on Pyramidal Successive Elimination", In : *International Computer Symposium*, pp. 41-44, Tainan, Taiwan, 1998.

[Mahb 02] A. MAHBOUBI, J. BENOIS-PINEAU, D. BARBA, "Tracking of objects in video scenes with time varying content", *EURASIP Journal of Applied Signal Processing, Special issue on Image Analysis for Multimedia Interactive Services*, Vol. 2002, N°6, pp. 582-594, June 2002.

[Marq 97] F. MARQUÈS, C. MOLINA, "Object Tracking for Content-Based Functionalities", In : *SPIE Visual Communications and Image Processing, VCIP'97*, pp. 190-199, San Jose, USA, 1997.

[Marq 98] F. MARQUÈS, J. LLACH, "Tracking of Generic Objects for Video Object Generation", In : *Proc. of the IEEE International Conference on Image Processing, ICIP'98*, pp. 628-632, Chicago, USA, 1998.

[Mont 91] A. MONTANVERT, P. MEER, A. ROSENFELD, "Hierarchical Image Analysis using Irregular Tessellations", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, N°4, pp. 307-316, April 1991.

[Pard 94] M. PARDÀS, P. SALEMBIER, *Joint Region and Motion Estimation with Morphological Tools*, pp. 93-100, *In J. Serra and P. Soille, editors, Mathematical Morphology and Its Applications to Image Processing*, Kluwer Academic Press, 1994.

[Park 00] D. PARK, H. YOON, C. WON, "Fast Object Tracking in Digital Video", *IEEE Transactions on Consumer Electronics*, Vol. 46, N°3, pp. 785-790, August 2000.

[Pate 00] S. PATEUX, "Tracking of video objects using a backward projection technique", In: *SPIE Visual Communications and Image Processing, VCIP'00*, pp. 1107-1114, Perth, Australia, 2000.

[Vinc 91] L. VINCENT, P. SOILLE, "Watersheds in Digital Spaces : an Efficient Algorithm Based on Immersion Simulations", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, N°6, pp. 583-598, June 1991.

[Wang 98] D. WANG, "Unsupervised Video Segmentation Based on Watersheds and Temporal Tracking", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 8, N°5, pp. 539-546, September 1998.

[Wu 95] L. WU, J. BENOIS-PINEAU, D. BARBA, "Spatio-Temporal Segmentation of Image Sequences for Object-Oriented Low Bit-Rate Image Coding", In : *Proc. of the IEEE International Conference on Image Processing, ICIP'95*, pp. 2406-2409, Washington DC, 1995.

Guillaume **Foret**

Diplômé ingénieur de l'Ecole Nationale Supérieure d'Électronique et de Radioélectricité de Grenoble (ENSERG-INPG) en 2000, Guillaume Foret a obtenu le titre de Docteur de l'Institut National Polytechnique de Grenoble, spécialité Signal Image Parole Télécoms, en 2003. Ses travaux de thèse ont porté sur la segmentation spatio-temporelle d'objets vidéo en vue de leur caractérisation. Il est actuellement attaché temporaire d'enseignement et de recherche à l'École Nationale Supérieure de l'Électronique et de ses Applications (ENSEA), Cergy-Pontoise. Il poursuit ses travaux de recherche au laboratoire ETIS (Equipes Traitement des Images et du Signal) dans le domaine de l'indexation et de la recherche d'images.

Pascal **Bertolino**

Après 6 années au service de sociétés privées, Pascal Bertolino est diplômé ingénieur CNAM en Génie Informatique en 1992 puis Docteur de l'Institut National Polytechnique de Grenoble, spécialité traitement d'images en 1995. Il est actuellement Maître de Conférences en informatique à l'IUT2 de l'Université Pierre Mendès France à Grenoble. Il effectue sa recherche au Laboratoire des Images et des Signaux sur le traitement et l'analyse des images et des vidéos.

Jean-Marc **Chassery**

Directeur de recherche au CNRS, développe son activité de recherche dans le domaine de l'analyse d'images appliquée à la représentation géométrique et au problème de la segmentation spatio-temporelle. Les modèles géométriques développés sont associés soit à la géométrie discrète soit à la géométrie à base de partitions par maillages.