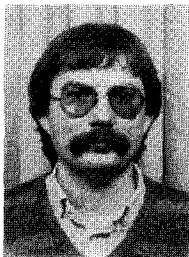


Estimation et structuration
d'indices spatio-temporels
pour l'analyse du mouvement
dans une séquence d'images

Motion segmentation and estimation through the determination
of spatio-temporal primitives in an image sequence



Patrick BOUTHEMY

IRISA/Centre INRIA, Campus de Beaulieu, 35042 RENNES CEDEX.

Ingénieur ENST, 1980. Thèse de Docteur-Ingénieur IRISA, Université de Rennes, 1982, en Traitement d'Image. Chargé de Recherche INRIA à l'IRISA depuis 1984. Activités de recherche en Analyse d'Image, plus particulièrement sur des problèmes liés à la prise en compte de séquences d'images, ainsi qu'en Compréhension d'Image.

RÉSUMÉ

Cet article présente un schéma d'estimation du mouvement dans des séquences d'images s'articulant en trois étapes fondamentales : *traitement local*, *structuration intermédiaire*, *estimation du champ des vitesses*. La première étape peut consister soit en l'obtention d'éléments de contour spatio-temporels avec leur direction spatiale et composante perpendiculaire de la vitesse, soit en une segmentation en régions de mouvement différent à partir d'une information partielle de mouvement. Les algorithmes proposés s'appuient tous les deux sur des principes de modélisation locale et de tests d'hypothèses mis en œuvre selon des méthodes de vraisemblance. Pour le premier point, une implantation de type convolution, de complexité équivalente à un simple détecteur de contour spatial, en sera exposée. La deuxième étape revient à structurer l'image en prenant en compte le chaînage des contours et/ou la partition en régions. Enfin, l'étape finale s'intéresse à l'estimation proprement dite du champ des vitesses, d'abord sur les contours obtenus, puis à l'intérieur des domaines qu'ils délimitent. Cette approche se caractérise notamment par une prise en compte explicite et préalable des discontinuités potentielles des champs de vitesses, dont l'estimation s'accompagne ainsi de l'obtention de zones délimitées et homogènes au sens du mouvement, ainsi que d'un ensemble d'indices spatio-temporels.

MOTS CLÉS

Séquence d'images, élément de contour en mouvement, région spatio-temporelle, champ de vitesses, segmentation, estimation, test de vraisemblance, gradient stochastique.

SUMMARY

This paper describes a motion estimation scheme from image sequences, which consists of three main stages: local processing, intermediate-level structuring, optic-flow field estimation. The first stage is concerned with either the determination of spatio-temporal edges along with their local spatial direction in the image plane and the component of their associated velocity vector perpendicular to this direction, or the segmentation into regions according to motion-based hierarchically-performed criteria which take into account an explicit partial motion information. Both designed algorithms utilize principles of local modeling and likely hypothesis testing. Concerning the first one, the defined likelihood ratio test is implemented according to some appropriate mask convolution, the complexity order of which is similar to conventional spatial gradient computation. The purpose of the second stage is to obtain a structured partition of the image, resulting from edge linking and/or region segmentation. The final stage deals with the velocity field estimation, that-is-to-say the reconstruction of the second component of displacement vectors by combining local observations. First, a recursive stochastic gradient, used to achieve the minimization of some simple functional, enables to estimate optic flow along contours. Then, the estimation within delineated domains is considered. Our approach is in particular distinguished by treating beforehand potential discontinuities of the velocity field in the image. Moreover, it provides with a set of intermediate-level spatio-temporal primitives.

KEY WORDS

Image sequence, moving edge, motion-based region segmentation, velocity field estimation, likelihood ratio, stochastic gradient.

1. Introduction

La considération de séquences d'images numériques ordonnées dans le temps introduit un champ spécifique d'investigation en analyse d'image, à savoir l'extraction d'informations de mouvement et plus généralement d'informations spatio-temporelles. Ce type d'études a émergé dans les années 70 à travers trois applications principales : le codage d'images de télévision et notamment la conception de codeurs avec compensation de mouvement [1-3], l'imagerie météorologique avec l'estimation de champs de vent via des traceurs nuageux [4], le suivi de cibles mobiles dans un contexte militaire [5].

Les techniques alors employées possédaient leurs caractéristiques propres liées aux contraintes et critères de l'application considérée, comme par exemple la qualité visuelle de l'image-intensité reconstruite, la limitation à des déplacements translationnels, ou la restriction à la présence d'un seul objet mobile, ... Aussi, les efforts se sont poursuivis ces dernières années, afin de concevoir ou compléter des schémas d'extraction d'informations liées au mouvement dans une séquence d'images [6-9].

De plus, de nouveaux domaines d'application sont apparus tels que la robotique pour des tâches d'évitement d'obstacles ou de navigation d'un robot mobile [10], le biomédical en ce qui concerne par exemple l'imagerie cardiaque [11], ou la surveillance du trafic routier. Le problème de l'estimation du mouvement est alors souvent abordé dans une perspective « analyse de scène », car il s'agit alors d'une information intrinsèque extraite des images permettant d'accéder à une connaissance de la scène 3D observée (profondeur et orientation des surfaces) et des mouvements 3D du capteur et/ou des objets dans la scène [7, 12-15].

Notre approche est à placer prioritairement dans un contexte algorithmique. Plus précisément, nous avons cherché à définir un ensemble de méthodes qui ne soit pas foncièrement dépendant d'une application ou

d'un type de scène particulier, mais qui puisse *a priori* prendre en compte un large éventail de cas. De plus, nous nous sommes attachés à construire un ensemble complet, qui s'attaque aussi bien aux aspects d'estimation que de segmentation, et qui puisse notamment traiter les configurations d'occlusion. Enfin, la formalisation présentée dans ce papier possède un caractère méthodologique général tout en pouvant se particulariser à de nombreux problèmes concrets d'estimation spatio-temporelle.

Nous nous plaçons donc dans un cadre général quant à la nature du problème traité. La caméra ou/et les objets observés peuvent être en mouvement; il n'y a pas *a priori* de restrictions quant au type de mouvement pris en compte (translation, rotation, homothétie, 2D ou 3D, ...); plusieurs types de déplacement peuvent être présents dans la même image. Le problème se pose alors en termes de robustesse. L'approche retenue, où segmentation et estimation sont conjointement abordées, permet de fournir un ensemble d'indices spatio-temporels de différents niveaux, exploitable dans une tâche d'analyse de scène 2D ou 3D. Selon l'application concernée, un sous-ensemble de ceux-ci peut suffire. L'information spatio-temporelle la plus riche pouvant être extraite est un champ structuré des vecteurs vitesses sur toute l'image. Soulignons que cet article sera principalement consacré à la détermination de primitives intermédiaires, entre les niveaux matrices de pixels et champ de vitesses apparentes sur toute l'image, et aux problèmes de segmentation et d'estimation qui lui sont liés.

La suite de cet article se structure de la façon suivante. Dans le prochain chapitre, nous présenterons le principe du schéma spatio-temporel à trois étapes que nous avons conçu. Puis nous détaillerons trois modules de ce schéma, à savoir la détermination locale d'éléments de contour spatio-temporels (chap. 3), la segmentation en régions selon un critère spatio-temporel (chap. 4), l'estimation du champ des vitesses le long d'une chaîne contour (chap. 5). Par ailleurs, une présentation de résultats sera associée à chaque chapitre.

2. Principe du schéma spatio-temporel à trois étapes

2.1. INTERACTION ENTRE ESTIMATION ET SEGMENTATION

Dans de nombreux problèmes d'estimation, une clef pour l'obtention de bonnes performances de l'estimateur est la réalisation préalable ou conjointe d'un processus de segmentation. Dans notre cas, une relative complexité supplémentaire provient de ce que la segmentation ne porte pas sur le signal d'entrée lui-même mais sur des quantités estimées, qui sont de plus bi-dimensionnelles, à savoir des vecteurs vitesses. Quelle que soit l'approche retenue, l'estimation d'un champ dense de vecteurs vitesses introduit, soit une prise en compte d'un certain voisinage, soit un processus itératif sur une zone de l'image [7, 9]. Il est alors postulé et utilisé que le champ des vitesses possède certaines propriétés de continuité. Sinon, seuls des cas restrictifs sont traités (tels que translation globale, ou considération de quelques éléments caractéristiques). Cependant, en général, le champ des vitesses apparentes, résultant du mouvement relatif de la caméra et des objets dans la scène observée, présente des sous-champs différents, et donc des discontinuités qu'il faut savoir gérer.

2.2. CONTEXTE BIBLIOGRAPHIQUE

Or, ce problème n'a pas été abordé initialement dans les schémas d'estimation, qu'il s'agisse de la classe des méthodes dites de mise en correspondance, comme la technique de relaxation avec des points de Moravec développée dans [16], ou de la classe des méthodes différentielles introduites dans [17]. Cela conduit alors à des erreurs d'estimation, comme il est noté dans [15]. Récemment, quelques schémas ont été proposés qui intègrent plus ou moins cet aspect de la question.

Dans [18], ce problème de discontinuités, liées en fait à des phénomènes d'« occlusion », est considéré. Le schéma proposé relève des méthodes de mise en correspondance. Il est basé sur la construction de pyramides d'images, sur une stratégie de « coarse-to-fine », et sur une fonction de similarité de type somme de carrés de différences. Cependant, le problème n'est pas véritablement levé, dans la mesure où seule une voie de résolution est indiquée à travers la définition de mesures de confiance et l'analyse de la géométrie locale de la surface d'appariement. Il en va sensiblement de même dans [19], où une formulation géométrique de l'équation différentielle de contrainte de mouvement définie dans [17] est donnée en termes de l'équation polaire d'une droite dans un espace de représentation approprié. Des mesures locales forment alors un « cluster » si elles correspondent à une zone cohérente de mouvement. Dans ces deux cas, il s'agit plus d'une détection éventuelle des discontinuités fondée sur une certaine fonction d'erreur et non d'une recherche explicite et formalisée de celles-ci.

En fait, la gestion de ces discontinuités potentielles, qui se présentent sur les frontières entre objets de mouvement différent ou à des profondeurs différentes (donc sur certains contours dans l'image), n'est véritablement

intégrée que dans le schéma d'estimation exposé dans [20]. Ce dernier découle dans son principe de celui de [17]. Cependant, une « métrique », construite à partir de dérivées premières et secondes de la fonction intensité, et traduisant le « motif » sous-jacent dans la zone considérée (contour, coin), est introduite dans la contrainte de lissage. Cette dernière est alors dénommée contrainte de lissage orienté. Ainsi, d'une certaine façon, la propagation de l'estimation n'est pas effectuée à travers les contours. Toutefois, cette approche ne permet pas de distinguer les véritables contours générateurs de discontinuités des contours intérieurs. Notons d'autre part que la résolution numérique résultante est relativement lourde.

2.3. APPROCHE DÉVELOPPÉE

Nous proposons une approche qui traite le problème de la segmentation de manière *explicite* et *préalable* à la phase finale d'estimation du champ complet des vitesses.

Nous avons formalisé le processus de détermination du mouvement apparent dans une séquence d'images selon trois étapes fondamentales :

1. *extraction de primitives locales;*
2. *structuration intermédiaire;*
3. *estimation du champ des vitesses.*

Cette méthodologie est détaillée dans le schéma-bloc présenté à la figure 1 où apparaissent les actions successives à mettre en œuvre et les niveaux de représentation impliqués correspondants. Une première version, où seul l'aspect contour était abordé, a été exposée dans [21]. Une alternative à cette approche *contour*, décrite au chapitre 3, a été également développée, à savoir une approche *région* qui sera exposée au chapitre 4. Le choix entre ces deux procédures peut résulter de l'application considérée, selon la nature des images traitées, les contraintes d'implantation, les informations utiles recherchées, ... Elles peuvent aussi être conjointement retenues; se pose alors le problème de la maintenance de la cohérence globale du processus, dont la prise en charge relève du module *Fusion*.

Tous les échelons principaux de ce schéma reposent sur une *modélisation* du sous-problème traité, ce qui présente le double intérêt suivant. Il s'agit de modèles stochastiques qui peuvent prendre en compte les cas les plus généraux. Toutefois, si une information *a priori* explicite est disponible, elle peut être correctement intégrée dans le modèle en question.

Comme nous l'avons souligné, un problème central pour l'estimation du champ des vitesses est de savoir gérer les discontinuités susceptibles d'y apparaître. Nous proposons donc de commencer par traiter ces discontinuités potentielles (segmentation spatio-temporelle). Ainsi, estimation et discrimination ne sont pas mêlées. La phase finale d'estimation du champ des vitesses s'effectue sur un domaine cohérent au sens du mouvement. De plus, ce schéma fournit des indices spatio-temporels intermédiaires (contours et régions avec leurs attributs spatio-temporels) déjà *exploitables*, qui peuvent d'ailleurs suffire dans certains cas.

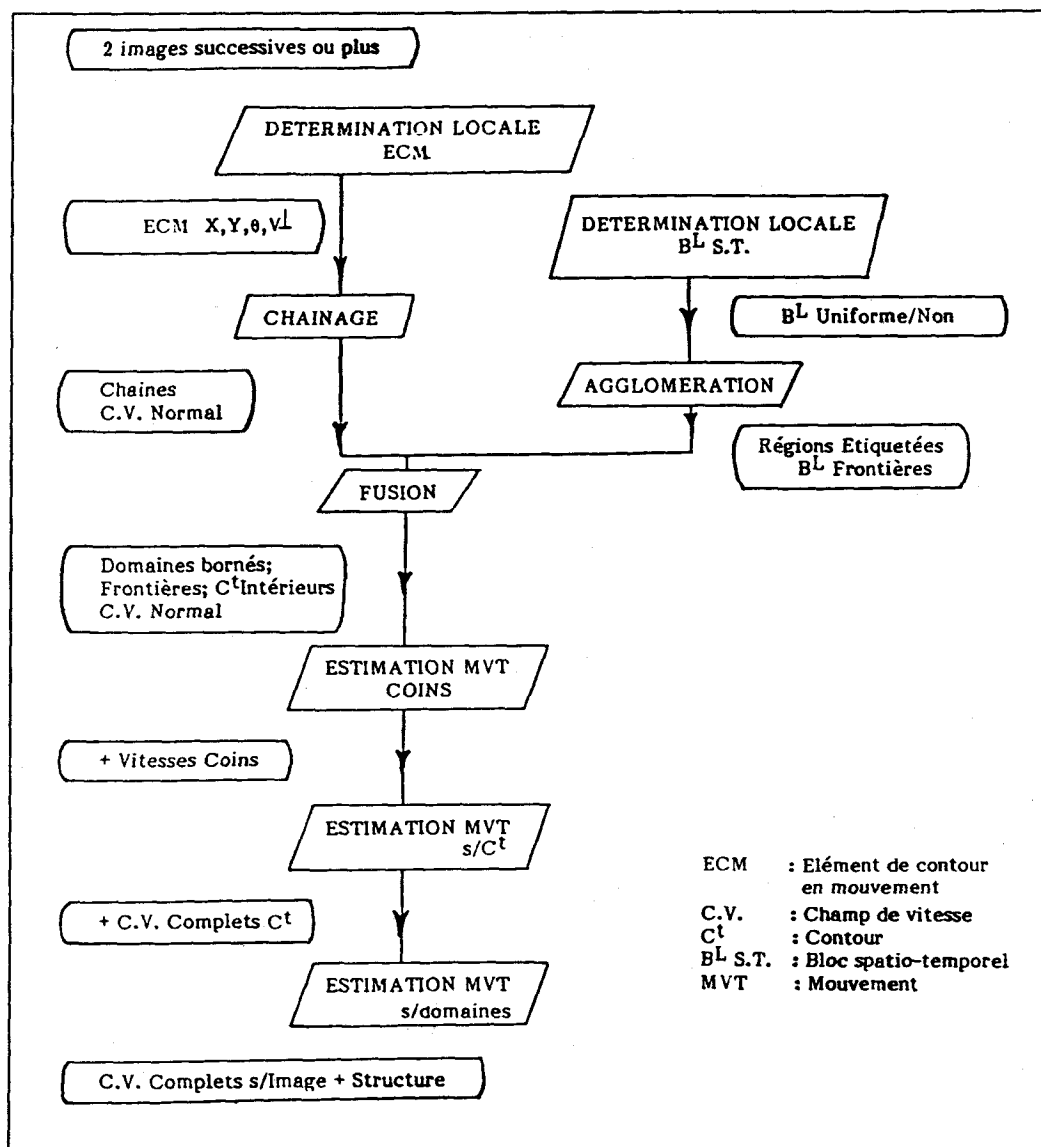


Fig. 1. — Schéma complet pour la détermination du mouvement apparent.

Par ailleurs, on aboutit à l'obtention d'un champ de vitesses *structuré*, c'est-à-dire partitionné en zones cohérentes au sens du mouvement, ce qui n'est pas le cas dans [20]. Les processus ultérieurs d'analyse et d'interprétation 3 D en sont alors facilités. Cela représente une alternative à la méthode à deux étapes d'hypothèse-et-vérification développée dans [15] pour la détermination du mouvement 3 D et de la structure d'objets multiples à partir du champ des vitesses apparentes dans une séquence d'images. Dans ce dernier cas, interprétation et segmentation se trouvent mêlées.

3. Détermination locale d'éléments de contour spatio-temporels

Ce premier module consiste en l'obtention d'éléments de contour spatio-temporels dont les attributs estimés sont :

1. leur position $\rho = (x_0, y_0)$ dans l'image;
2. leur direction spatiale locale θ dans l'image;

3. la composante du vecteur vitesse associé v^\perp perpendiculaire au contour.

Il est bien connu qu'on ne peut localement (c'est-à-dire à travers une « fenêtre d'observation locale ») estimer la composante tangentielle au contour du vecteur vitesse, ou en d'autres termes la composante du glissement du contour sur lui-même [22, 23]. Un premier traitement local est toutefois nécessaire si l'on vise l'obtention d'un champ dense de vecteurs vitesses. L'estimation de la deuxième composante du vecteur vitesse doit donc être repoussée au niveau suivant du processus, via une combinaison adéquate d'observations locales.

3.1. MODÉLISATION D'UN ÉLÉMENT DE CONTOUR EN MOUVEMENT

Une séquence d'images est considérée comme étant un espace à trois dimensions (x, y, t) , deux dimensions spatiales et une dimension temporelle. Dans cet espace, un élément de contour en mouvement ou

élément de contour spatio-temporel, que l'on notera pour simplifier ECM, génère une portion de surface. Différentes modélisations de cette surface plus ou moins complexes peuvent être envisagées.

3.1.1. Discussion de l'approche différentielle

A défaut de toute information complémentaire, il est en fait censé de choisir la plus simple et la plus tractable : une portion de surface planaire, d'autant qu'il s'agit d'un traitement local. Ce choix s'apparente à l'approximation du premier ordre introduite dans les méthodes différentielles [17, 22-24], qui conduit à l'équation de base reliant le vecteur vitesse dans l'image $v=(dx/dt, dy/dt)$ au gradient spatial de la fonction intensité f (ou d'une version filtrée), $\nabla f=(\partial f/\partial x, \partial f/\partial y)$ et à la dérivée temporelle de f , $\partial f/\partial t$, selon :

$$(1) \quad v \cdot \nabla f = -\partial f/\partial t.$$

(On peut remarquer que l'on obtient bien analytiquement par cette équation la seule composante du vecteur vitesse colinéaire au gradient spatial, c'est-à-dire perpendiculaire à l'élément local de contour.)

Nous avons cherché à concevoir une méthode qui puisse intrinsèquement satisfaire à deux critères, que ne remplit pas l'équation (1), à savoir :

1. la prise en compte de n'importe quel type de contour, y compris les contours d'« occlusion »;
2. la possibilité de mesurer une amplitude quelconque de déplacement.

En effet, les conditions sous-jacentes liées à la différentiation ne sont en fait pas satisfaites pour les contours d'« occlusion » (pour les contours intérieurs, un filtrage passe-bas préalable permet de mieux y répondre). Par ailleurs, l'approche différentielle convient pour de petits déplacements, ou plus précisément tant que la relation décrite par l'équation (1) a un sens, c'est-à-dire tant que l'amplitude de la composante normale de la vitesse ne dépasse pas l'étendue en pixels de la pente de la fonction intensité (étendue pouvant être « allongée » après filtrage passe-bas). Notons aussi que l'équation (1) suppose que la brillance du point correspondant dans l'espace ne varie pas dans l'intervalle de temps considéré.

3.1.2. Configurations et hypothèses associées

Nous considérons donc le cas d'une portion de plan comme modèle d'ECM. Cette portion de plan est engendrée dans l'espace (x, y, t) :

1. par l'élément de contour spatial référencé au point x_0, y_0 et représenté par un petit segment de droite dans un plan $t=t_0$;
2. par son vecteur vitesse associé $V=(dx/dt, dy/dt, 1)=(v, 1)$.

En pratique, puisqu'il s'agit de géométrie discrète, on fera la confusion entre vecteur vitesse et vecteur déplacement (au facteur dt près).

Cette modélisation est illustrée par la figure 2. L'orientation de la portion de plan est définie par deux angles, cf. figure 2 :

1. l'angle θ repéré par rapport à l'axe des x dans un plan $t=t_0$, ou direction spatiale locale de l'élément de contour;

2. l'angle ψ repéré par rapport à l'axe des t .

Il est évident que la seule détermination de cette portion de plan ne permet d'accéder qu'à la seule composante normale au contour de la vitesse, donnée par $v^+ = \text{tg } \psi$.

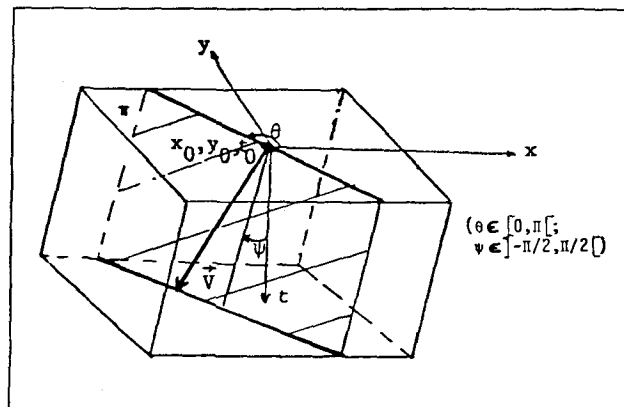


Fig. 2. — Modélisation locale planaire d'un ECM.

Le paragraphe suivant exposera la méthode définie permettant la détection directe de telles portions de plan et l'estimation simultanée de leurs paramètres (θ, ψ) . Elle est basée sur la maximisation d'un rapport de vraisemblance généralisée. Soulignons qu'il s'agit bien par cette approche de la détermination des ECMs, c'est-à-dire simultanément détection et estimation des paramètres, et non seulement de la détection des ECMs comme dans [8], où un opérateur de gradient spatial est multiplié avec une différence temporelle, et dans [24] où l'on considère les passages par zéro du d'Alembertien de la séquence d'images convoluée avec un Gaussien tri-dimensionnel en (x, y, t) .

Au préalable, il est nécessaire de préciser les deux hypothèses de configuration locale intervenant dans le critère. Étant donné un volume élémentaire π placé au point p dans l'espace (x, y, t) , deux hypothèses peuvent intervenir :

– hypothèse H_0 : il n'existe pas d'ECM dans π ; alors, la distribution des intensités au sein de π est modélisée selon niveau constant $c_0 + \text{bruit}$, où le bruit est un bruit Gaussien centré de variance σ^2 .

– hypothèse H_1 : il existe un ECM dans π ; alors, cela se traduit par la présence d'une portion de plan S divisant π en deux sous-volumes π_1 et π_2 où les niveaux constants des modèles d'intensité, du même type que précédemment, sont respectivement c_1 et c_2 , avec $c_1 \neq c_2$.

Notons que le cas d'un contour statique relève bien de l'hypothèse H_1 , faisant partie des contours spatio-temporels. On a alors $V=(0, 0, 1)$ et $\psi=0$. D'un point de vue de terminologie, les ECMs incluent les contours statiques, c'est-à-dire de déplacement nul.

3.2. TEST DE VRAISEMBLANCE POUR LA DÉTERMINATION DES ECMs

A chacune des hypothèses précédemment définies est associée une fonction de vraisemblance, respectivement L_0 et L_1 . Cette fonction est en fait la densité jointe des lois de probabilité des intensités aux points

du volume π . D'après ce qui précède, ces dernières suivent des lois gaussiennes, à savoir :

- $N(c_0, \sigma^2)$ en tout point de π pour l'hypothèse H_0 ;
- $N(c_1, \sigma^2)$, resp. $N(c_2, \sigma^2)$, en tout point de π_1 , resp. π_2 , pour l'hypothèse H_1 .

Si l'on suppose les variables indépendantes, chaque fonction de vraisemblance s'exprime comme le produit des densités des lois considérées.

Le choix de l'hypothèse H_1 contre l'hypothèse H_0 est basé sur l'optimisation, en fonction du vecteur de paramètres $\Lambda = (\rho, \theta, \psi, c_0, c_1, c_2)$ du rapport logarithmique ξ des deux fonctions L_1 et L_0 . Notons que la variance σ^2 n'est pas comprise dans le vecteur Λ . Elle est supposée identique d'une hypothèse à l'autre, c'est-à-dire que l'on considère que la variance du bruit n'est pas « liée » au niveau du signal. Elle est donc supposée constante au regard du test de vraisemblance. L'hypothèse H_1 est sélectionnée si la valeur maximale du rapport $\xi(\Lambda)$ obtenue pour les paramètres optimaux $\hat{\Lambda}$ est supérieure à un seuil prédéterminé λ_{ecm} ; sinon, c'est l'hypothèse H_0 qui est retenue. Ce test peut se représenter par la formule suivante :

$$(2) \quad \max_{\rho, \theta, \psi} \max_{c_1, c_2} \min_{c_0} \xi(\Lambda) \geq \lambda_{ecm}$$

Les paramètres qui interviennent dans le critère de décision sont en fait de deux ordres. D'une part, on peut distinguer les paramètres que l'on peut qualifier de paramètres géométriques qui déterminent la portion de plan modélisant l'ECM. D'autre part, apparaissent les paramètres dits radiatifs, c'est-à-dire les différentes moyennes qui interviennent dans les distributions d'intensité.

Or, on ne peut dériver, comme il est montré dans [21] et [25], une formulation analytique explicite des paramètres géométriques optimaux. Aussi, nous allons considérer un jeu prédéfini de G géométries données $S_j, j=1, \dots, G$, dans le domaine de variation possible des paramètres associés $\Phi = (\theta, \psi)$; $G = G_\theta \cdot G_\psi$. Par contre, à géométrie donnée S_j , c'est-à-dire à $\Phi_j = (\theta_{j_k}, \psi_{j_l})$, on peut dériver l'expression analytique exacte des paramètres radiatifs optimaux; $\hat{c}_0, \hat{c}_1, \hat{c}_2$, qui sont en fait les moyennes empiriques calculées sur les volumes correspondants. Le rapport de vraisemblance ξ peut alors se mettre sous la forme suivante [21],

$$(3) \quad \xi(\rho, \Phi_j) = \frac{n_1 n_2}{2(n_1 + n_2) \sigma^2} (\hat{c}_1 - \hat{c}_2)^2$$

Si l'hypothèse H_1 est sélectionnée, on peut alors conclure qu'est présent au point $\hat{\rho}$, un ECM dont les paramètres estimés sont précisément donnés par les valeurs des paramètres optimaux qui maximisent le rapport ξ , à savoir pour direction spatiale $\hat{\theta}_{j_k}$, et pour composante de vitesse normale

$$v^\perp = \text{tg } \hat{\psi}_{j_l}$$

3. 3. MISE EN ŒUVRE DU PROCÉDÉ

Pour une géométrie donnée S_j au point ρ , le calcul du rapport logarithmique de vraisemblance, $\xi(\rho, \Phi_j)$, s'apparente en fait, après quelques développements

mathématiques, à un produit de convolution de la fonction intensité avec un noyau ou masque tridimensionnel, dont les coefficients ne dépendent que de la géométrie S_j [21, 25]. Par conséquent, se donner un jeu de géométries revient à calculer hors-ligne le jeu de masques associés.

Plus formellement, on considère la racine carrée de $\xi(\rho, \Phi_j)$, notée $\zeta(\rho, \Phi_j)$, qui s'écrit de la façon suivante :

$$(4) \quad \zeta(\rho, \Phi_j) = \left| \sum_{m \in M} a_j(m) \cdot f(\rho + m) \right|$$

ou

$$(5) \quad \zeta(\rho, \Phi_j) = \left| \sum_{t=t_1}^{t_2} \sum_{m_t \in M_t} a_j(m_t) \cdot f(\rho + m_t) \right|$$

avec

$$(6) \quad \left\{ \begin{array}{l} \text{(i) } a_j(m_t) = \frac{v_t}{\sigma} \left(\frac{n_2}{2n_1(n_1+n_2)} \right)^{1/2}, \\ \quad \text{si } m_t \in \pi_1; \\ \text{(ii) } a_j(m_t) = -\frac{v_t}{\sigma} \left(\frac{n_1}{2n_2(n_1+n_2)} \right)^{1/2}, \\ \quad \text{si } m_t \in \pi_2. \end{array} \right.$$

où $f(\cdot)$ note les valeurs observées de l'intensité, $M = \cup M_t$ représente un jeu d'indices à trois dimensions tels que $\rho + m$ désigne tour-à-tour tous les points du volume élémentaire $\pi = \pi_1 \cup \pi_2$ placé au point ρ ; n_1 , resp. n_2 , le nombre de points de π_1 , resp. π_2 . Notons que le procédé autorise directement la prise en compte aussi bien de deux images successives t_1, t_2 , que de plus de deux images, $t_1, t_1+1, \dots, t_2-1, t_2$. Enfin, v_t désigne un coefficient compris entre 0 et 1, tenant compte de l'aspect discret, à savoir si le segment de droite, intersection de la portion de plan S_j et du plan image t , passe par le point m_t , $v_t=0$, traverse la « maille » correspondante, $0 < v_t < 1$, ou n'y passe pas, $v_t=1$. Précisons que v_t est donné par la relation suivante $v_t = |\delta_1 - \delta_2| / (\delta_1 + \delta_2)$, où δ_1 , resp. δ_2 , représente l'aire de la « maille » incluse dans π_1 , resp. π_2 , et dans le plan image t .

Une première version de ce procédé [21, 25], calculait, aux points considérés de l'image, tous les $\zeta_j := \zeta(\rho, \Phi_j)$, c'est-à-dire effectuait les convolutions avec tous les masques $\{a_j\}$, puis retenait la sortie maximale. Or, si l'on considère entre deux images successives des déplacements perpendiculaires v_j^\perp aboutissant à des positions « entières » de pixels, le facteur ne dépend plus de t , $v_t = v$. Ceci n'est en fait pas du tout restrictif, et est même le cas standard si l'on ne considère que deux images; l'estimation sous-pixel se devant d'être rejetée hors convolution à l'aide d'une interpolation locale sur la courbe de vraisemblance dans la direction retenue.

Une nouvelle version, de complexité équivalente à un simple détecteur de contour spatial de type gradient, en a été posée. Le principe en est le suivant.

1. Chaque image t de la séquence est convoluée avec un jeu de G_θ masques A_{j_k} bidimensionnels, dont les coefficients ne dépendent plus de t , mais en fait seule-

ment de l'aspect spatial de la géométrie du modèle dans le plan image de référence $t=t_0$. (Rappelons que G_0 correspond au nombre de directions spatiales quantifiées possibles pour l'élément de contour.) Ces coefficients sont fournis par les expressions (6) (i)-(ii).

2. En chaque point de l'image t_0 , le calcul du rapport ζ_j pour chaque géométrie spatio-temporelle S_j considérée, revient alors à la sommation des deux (ou plus) valeurs filtrées avec le masque correspondant, prises en ce point et en une ou des positions adéquates (c'est-à-dire à $v_j^\perp \Delta t$) de ce point dans la direction orthogonale à θ_{jk} dans l'autre ou les autres images considérées, suivie de la prise de la valeur absolue de cette somme.

3. La valeur ζ_j maximale est retenue et l'ECM correspondant est enregistré ainsi que ses paramètres si cette valeur est supérieure au seuil λ_{ecm} .

En plus du caractère rapide de cette version, les avantages suivants sont obtenus : non-limitation *a priori* de l'espace de recherche via le choix préalable de G_ψ déplacements normaux quantifiés; dépendance négligeable du coût calcul vis-à-vis du nombre de déplacements effectivement considérés.

Enfin, il n'a pas été évoqué la maximisation de ζ_j en fonction de ρ . Elle est en pratique implantée de manière analogue à une procédure d'affinage.

Par ailleurs, concrétisons la remarque antérieure concernant la possibilité de mesurer une amplitude quelconque de déplacement, tout en maintenant le caractère local du traitement, en termes de voisinage spatial dans chaque image t et de nombre d'images successives utilisées. En effet $\psi \in]-\pi/2, \pi/2[$ implique que $v^\perp \in]-\infty, \infty[$. Bien sûr, ceci doit s'entendre intrinsèquement. En pratique, cette propriété sera relativisée selon la nature des images considérées, en particulier la densité spatiale réelle des contours.

Il faut souligner aussi que le formalisme lié à la détermination d'un ECM et sa mise en œuvre par convolution permettent de façon simple et immédiate l'introduction de modélisations plus complexes (telles qu'une portion de cylindre, si l'on modélise l'élément de contour spatial par un arc de cercle, ou un autre type de surface réglée si tout en conservant la modélisation de l'élément de contour par un segment de droite on ajoute une rotation de l'élément de contour au déplacement local, ...). Il s'agira simplement de construire hors-ligne le jeu de masques (c'est-à-dire calculer un ensemble de coefficients) correspondant à ces nouvelles géométries. Par contre, l'implantation rapide décrite ci-dessus ne sera pas forcément possible; il faudra éventuellement avoir alors recours à la version présentée en [21] et [25].

3.4. RÉSULTATS

Détaillons quelques éléments supplémentaires dans l'implantation de l'algorithme. Un ECM est évidemment préalablement un élément de contour spatial. Aussi, une détection de contours spatiaux pourrait être effectuée dans un premier temps, comme dans [23], suivie de l'estimation pour ces points contours de la composante perpendiculaire de la vitesse. Cependant, il peut être mis en évidence que la méthode proposée prend également en compte cette remarque.

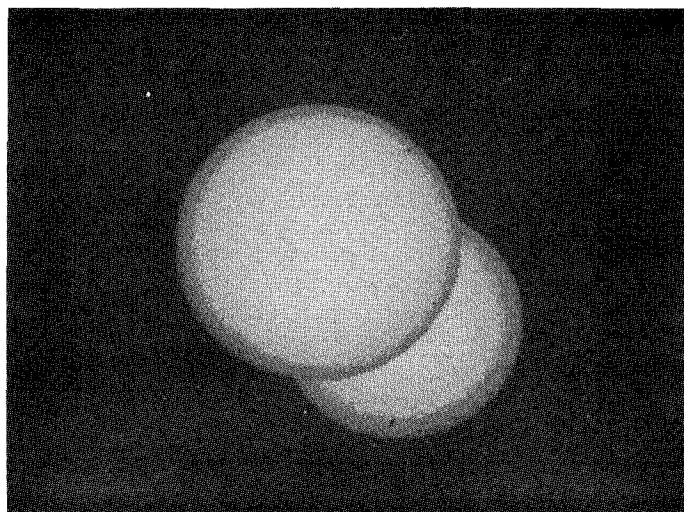
En effet, l'expression (5) peut se réécrire, en posant

$$\zeta_t := \zeta_t(\rho, \Phi_j) := \sum_{m_t \in M_t} a_j(m_t).f(\rho + m_t),$$

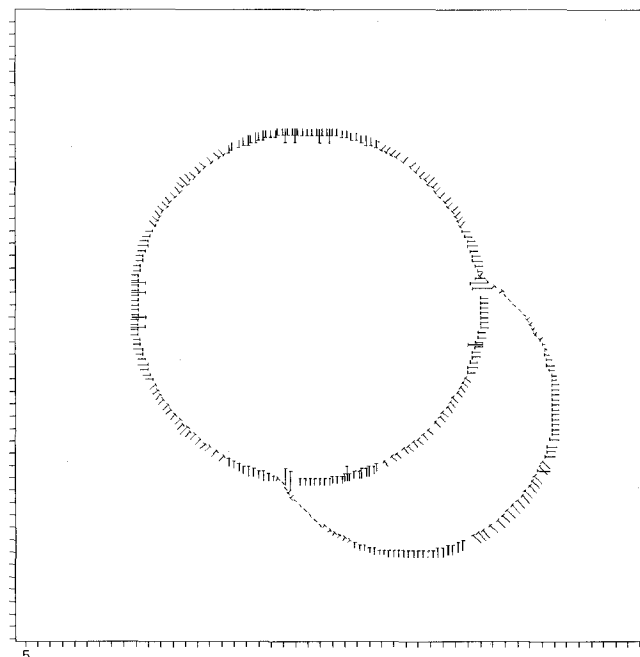
comme suit :

$$\zeta_j = \zeta(\rho, \Phi_j) = \left| \zeta_{t_0}(\rho, \theta_{jk}) + \sum_{t=t_1}^{t_2 \setminus \{t_0\}} \zeta_t(\rho, \theta_{jk}, v_{jt}^\perp) \right|$$

où t_0 représente le plan image auquel sont affectés les ECMs cherchés; ζ_{t_0} ne dépend pas de v_{jt}^\perp . On peut alors imposer qu'on ne considère le calcul complet de ζ_j que si $|\zeta_{t_0}| > \eta \lambda_{ecm}$, où $0 < \eta < 1$. Soulignons que la décision finale exprimée par le critère (2) reste spatio-temporelle. (Par ailleurs, afin d'augmenter la robustesse du procédé, on peut éventuellement intro-



(a)

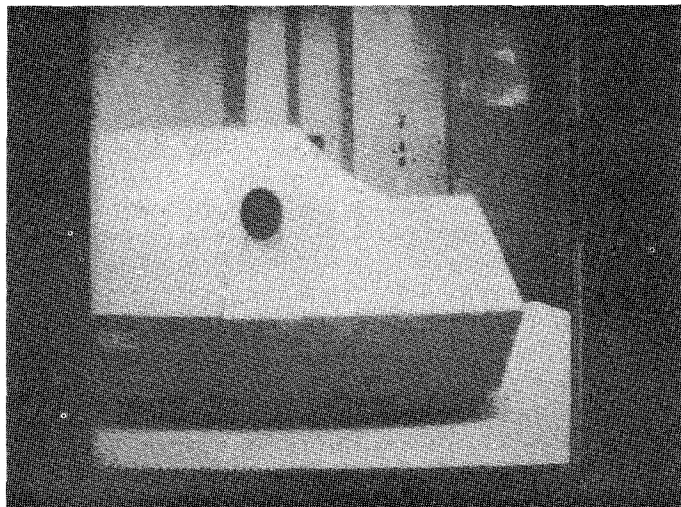


(b)

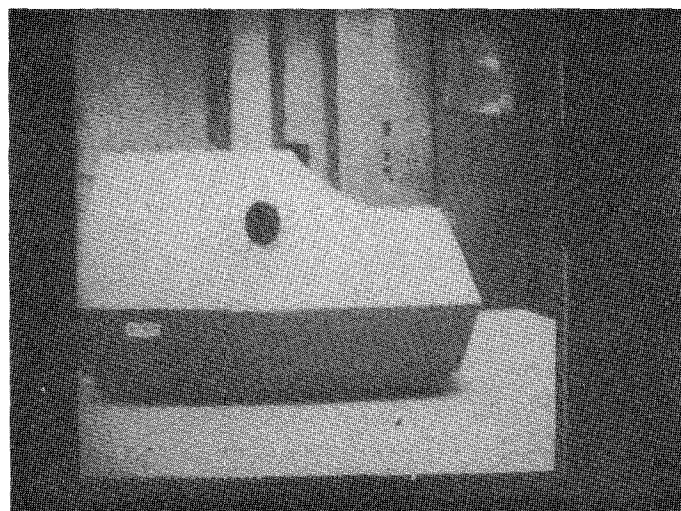
Fig. R1. — (a) Première des deux images traitées. (b) ECMs déterminés avec leur composante perpendiculaire de déplacement (un ECM sur deux est tracé).

duire une contrainte supplémentaire, à savoir $\tau_1 \leq |\zeta_t/\zeta_{t_0}| \leq \tau_2$, pour $t_1 \leq t \leq t_2$ et $t \neq t_0$. Ces ajouts résultent du fait que la procédure d'affinage est une implantation approchée de la maximisation de ζ en fonction de ρ .)

Deux exemples de résultats sont présentés. Le premier concerne des images générées par ordinateur. Elles comprennent deux disques dont la fonction d'intensité est sphérique (par analogie avec l'équation d'une surface sphérique) sur un fond uniforme bruité, le premier se dilatant de trois pixels et recouvrant partiellement le second qui se déplace diagonalement vers le coin inférieur droit de (3, 3) dans le plan image (voir fig. R1a). On a donc le cas d'une frontière d'« occlusion » entre deux objets en mouvement. Les ECMs détectés sont montrés avec leur composante perpendiculaire de vitesse à la figure R1b. La contrainte supplémentaire « des rapports respectifs » n'est pas prise en compte (ou cela revient à $\tau_1 = 0$ et $\tau_2 = \infty$). Le seuil λ_{ecm} a été fixé à 2 500 et η à 0,5. Les masques A_j sont de taille 5×5 et six orientations θ_{jk} (0, 30, 60, 90, 120, 150) ont été considérées.



(a)



(b)

Fig. R2. — (a) Première image « imprimante ».
(b) Seconde image « imprimante ». Taille : 128×128 .

Le deuxième exemple comporte des images réelles acquises au laboratoire avec une caméra CCD. L'imprimante a été déplacée à peu près dans l'axe de la caméra, cette dernière comme le fond restant fixe, figure R2a-b. Les masques utilisés sont à nouveau de taille 5×5 , quatre orientations θ_{jk} ont été choisies (0, 45, 90, 135), $\lambda_{ecm} = 5\,000$ ($\tau_1 = 0,75$ et $\tau_2 = 1,25$). Les résultats sont représentés à la figure R3.

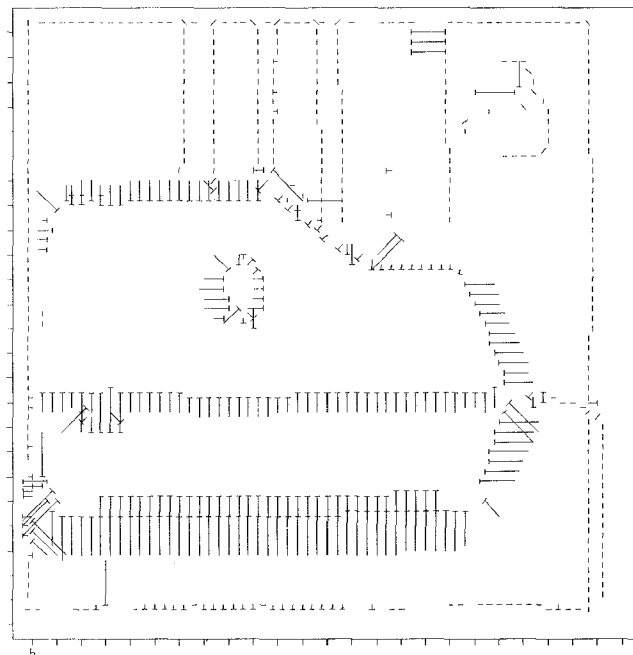


Fig. R3. — ECMs déterminés pour l'exemple de la figure R2.

Précisons que le module d'estimation du déplacement des coins permettra d'améliorer le traitement des coins, qui peut être insuffisant à ce niveau local. En effet, la recherche ne s'effectuant que dans la direction perpendiculaire à l'élément de contour, de tels « vis-à-vis » peuvent ne pas être présents d'une image à l'autre dans le cas des coins. Ce module ne sera qu'une extension immédiate de l'approche qui vient d'être décrite pour la détermination des ECMs, le modèle géométrique devant simplement être plus complexe. On obtiendra d'ailleurs directement le vecteur vitesse complet pour les coins.

4. Segmentation en régions selon un critère de mouvement

Nous avons étudié un procédé de segmentation en régions selon un critère spatio-temporel fondé sur une information partielle de mouvement. Ce module peut intervenir de deux façons. Il pourrait être la seule étape initiale dans un processus de détermination du mouvement apparent, induisant une partition de l'image en zones cohérentes spatio-temporellement. Il pourrait même suffire si le but poursuivi est de différencier les zones fixes des zones mobiles et de distinguer les différentes zones mobiles. D'autre part, il peut être conçu comme fonctionnant en parallèle avec le module réalisant la détermination de contours

spatio-temporels. A l'issue des deux traitements, un processus coopérant pourrait améliorer les deux lots de primitives (cf. module *Fusion* de la figure 1), en particulier éliminer les estimations erronées, permettre d'obtenir des contours fermés (là où le chaînage de contours s'avère insuffisant), trier entre contours frontières et contours intérieurs, ...

Les premiers schémas de segmentation spatio-temporelle s'intéressaient uniquement à séparer les régions où un changement temporel s'est produit dans l'image, des régions fixes [8, 26, 27]. Récemment, des techniques de segmentation de champs de vitesses ont été proposées, soit par recherche des frontières, via l'application d'un laplacien sur les composantes des vecteurs vitesses [28], soit par identification aux moindres carrés d'un modèle de vitesses à travers une structure pyramidale [29]. Cependant, comme nous l'avons souligné, il nous semble hautement préférable d'aborder le problème de la segmentation avant toute estimation complète du champ des vitesses, plutôt que d'effectuer la démarche inverse. Nous avons donc défini un procédé de segmentation en zones cohérentes au sens du mouvement à partir d'une information *partielle* de mouvement. Plusieurs modèles de vitesse, et donc de critères d'homogénéité associés, peuvent intervenir de manière hiérarchique à travers le même formalisme de vraisemblance.

4. 1. CRITÈRE AVEC MODÈLE DE VITESSE CONSTANTE

Nous associons, dans un premier temps, à une région cohérente spatio-temporellement le modèle de vitesse suivant :

$$\text{constant } \mathbf{T} + \text{bruit}$$

A partir de ce modèle de vitesse, nous allons construire le critère d'homogénéité pour une région donnée. Comme il l'a été déjà précisé, nous disposons en fait d'une observation partielle de la vitesse \mathbf{v} au point (x, y) , à savoir $\mathbf{v}(x, y) \cdot \nabla f(x, y)$. Considérons l'expression suivante :

$$(7) \quad e(x, y) = [\mathbf{T} - \mathbf{v}(x, y)] \cdot \nabla f(x, y)$$

On pose que e est une variable aléatoire suivant une loi gaussienne centrée de variance σ^2 et paramétrée par \mathbf{T} . Si $\mathbf{T} = (a, b)$, notons cette variable $e_{a,b}$.

Afin d'explicitier e , on a recours à l'équation exprimée en (1), à savoir $\mathbf{v} \cdot \nabla f = -\partial f / \partial t$. Dans ce contexte-ci d'utilisation, certaines des remarques formulées à l'encontre de cette équation ne sont plus gênantes.

En effet là où cette équation n'est plus valable (contours d'« occlusion ») et où alors le critère a toutes les chances de diverger, se trouvent précisément les ruptures que l'on veut détecter. Parallèlement, là où l'information de gradient est pauvre (zone quasi homogène spatialement) et où donc cette équation est peu opérante, le critère « tendra vers 0 », mais la zone considérée a toutes les chances d'être aussi cohérente temporellement.

e s'exprimera donc de la façon suivante, après explicitation dans (7) des vecteurs par leurs composantes et prise en compte de la relation (1) :

$$(8) \quad e_{a,b} = a \cdot \partial f / \partial x + b \cdot \partial f / \partial y + \partial f / \partial t$$

Le critère s'appuie sur un test de rapport de vraisemblance. Étant donnée une zone Z de l'image, deux hypothèses sont considérées :

- hypothèse H_0 : en tout point de Z , $e := e_{a_0, b_0}$;
- hypothèse H_1 : Z se subdivise en deux sous-zones Z_1 et Z_2 telles que, pour tout point de Z_1 , $e := e_{a_1, b_1}$, pour tout point de Z_2 , $e := e_{a_2, b_2}$; avec $(a_1, a_2) \neq (a_2, b_2)$.

A chaque hypothèse est associée une fonction de vraisemblance, produit des densités des lois gaussiennes de e , supposées indépendantes d'un point à l'autre, respectivement L_0 et L_1 . Si l'on note χ le rapport logarithmique de L_1 sur L_0 , il s'écrit comme suit :

$$(9) \quad \chi(\Omega) = \frac{1}{2\sigma^2} \sum_{(x,y) \in Z} e_{a_0, b_0}(x, y)^2 - \frac{1}{2\sigma^2} \sum_{(x,y) \in Z_1} e_{a_1, b_1}(x, y)^2 - \frac{1}{2\sigma^2} \sum_{(x,y) \in Z_2} e_{a_2, b_2}(x, y)^2$$

avec $\Omega = (a_0, b_0, a_1, b_1, a_2, b_2)$. Les paramètres optimaux $\hat{\Omega}$ maximisant les fonctions de vraisemblance sont obtenus en résolvant les systèmes linéaires suivants :

$$\begin{cases} \frac{\partial \chi}{\partial a_i} = 0 \\ \frac{\partial \chi}{\partial b_i} = 0 \end{cases}$$

Soit λ_{reg} un seuil prédéterminé. L'hypothèse H_0 l'emporte et la région Z est donc décidée homogène au sens du mouvement, si

$$(10) \quad \chi(\hat{\Omega}) \leq \lambda_{reg}$$

Les détails de ces calculs peuvent être trouvés dans [30]. En fait, on aboutit à l'expression (9) pour le rapport χ , car il est posé que la variance de la variable e ne dépend pas de l'hypothèse considérée et est supposée connue. En toute rigueur, la variance de e peut être formulée selon $\sigma^2 = a^2 \sigma_x^2 + b^2 \sigma_y^2 + \sigma_t^2$ où σ_x^2 , σ_y^2 et σ_t^2 sont respectivement les variances de $\partial f / \partial x$, $\partial f / \partial y$ et $\partial f / \partial t$, prises comme variables aléatoires indépendantes. Cela conduirait alors à un système d'équations non linéaires en a et b pour estimer les valeurs optimales de ces paramètres, et donc à une résolution non immédiate. Négliger ce développement permet d'obtenir un critère directement manipulable. Cette simplification n'est de plus pas aberrante dans un contexte de segmentation, où il est fréquent qu'une méthode soit tout aussi robuste si un modèle moins « fin » est utilisé; en l'occurrence, il n'est pas indispensable d'introduire l'estimation de la variance de la variable observée (ou alors *a posteriori*).

4. 2. PROCESSUS DE SEGMENTATION

La stratégie employée est du type division-et-agglomération [31]. On considère deux images successives. Dans une première phase, on établit une subdivi-

vision de l'image de référence en blocs 16×16 . Il n'est pas question, dans cette phase, d'identifier la géométrie des sous-zones éventuelles Z_1 et Z_2 d'un bloc ou zone donnée Z . Aussi, nous allons appliquer le critère défini au paragraphe précédent pour deux configurations préétablies. Ces deux configurations, que l'on nommera horizontale et verticale, sont présentées à la figure 3.

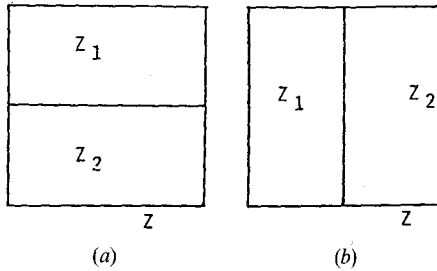


Fig. 3. — Configurations de subdivision préétablies, (a) horizontale, (b) verticale.

Deux rapports de vraisemblance, correspondant aux deux configurations considérées, sont alors calculés en parallèle, χ_H et χ_V , avec les paramètres optimaux adéquats. Le rapport retenu pour le critère de décision (10) est $\chi = \max(\chi_H, \chi_V)$. Pour les zones classées non homogènes, une nouvelle subdivision en blocs 4×4 est effectuée. Le même critère est à nouveau appliqué. On obtient ainsi un ensemble de blocs 16×16 ou 4×4 classés homogènes et de blocs 4×4 non homogènes, que nous appellerons aussi blocs-frontière. Il va de soi que, selon les cas, d'autres tailles de blocs pourraient être considérées ainsi qu'un nombre supérieur d'itérations de subdivision.

Intervient ensuite la phase itérative d'agglomération. Elle repose sur le même critère (10) que précédemment. Les deux régions candidates sont alors les sous-zones Z_1 et Z_2 , la zone Z étant leur union potentielle. Cette phase est menée de façon plutôt « parallélisée » en ce sens que les régions sont agglomérées peu à peu dans toute l'image. Plus précisément, un couple de deux régions est pris en considération si la taille de leur union est inférieure à un seuil qui augmente à chaque itération. Ce mode de fusion a été retenu en raison du type de critère utilisé. Le processus s'arrête lorsque, pour chaque région, aucune agglomération supplémentaire n'est possible avec les régions de sa liste d'adjacence.

On aboutit ainsi à un ensemble de régions étiquetées, plus éventuellement une collection de blocs élémentaires 4×4 frontières, initiaux (c'est-à-dire obtenus à l'issue de la phase 1 de division) si ces derniers n'ont pas été pris en compte dans la phase d'agglomération, résiduels dans le cas contraire. Notons que ces blocs élémentaires sont situés normalement sur des zones recouvertes ou découvertes par un objet d'une image à l'autre. Il pourrait d'ailleurs être intéressant pour certaines applications de les traiter à part en tant que tels.

4. 3. CRITÈRE AVEC MODÈLE DE VITESSE PLUS ÉLABORÉ

Pour les objets présentant un champ de vitesses apparentes dans l'image d'un type éloigné du modèle

retenu, une sur-segmentation peut apparaître. On procède alors, pour les régions de mouvement apparent non nul, à une deuxième étape de fusion à l'aide d'un modèle plus riche. Le modèle considéré est du type linéaire, c'est-à-dire « *variation de vitesse constante + bruit* ».

Soit (x_g, y_g) , le centre de gravité d'une région R telle que $(a, b) \neq (0, 0)$. [Précisons que ce vecteur représentatif de la cohérence de R au sens du mouvement est en fait le vecteur (\hat{a}_0, \hat{b}_0) estimé lors de la constitution de R .] Pour tout point (x, y) de R , écrivons le développement au premier ordre du vecteur vitesse :

$$\mathbf{v}(x, y) = \mathbf{v}(x_g, y_g) + (x - x_g) \partial \mathbf{v} / \partial x + (y - y_g) \partial \mathbf{v} / \partial y + O^2$$

Le nouveau critère de fusion de deux régions repose sur le modèle de gradient de vitesse suivant :

$$\text{constant} (\mathbf{Dc}, \mathbf{Dl}) + \text{bruit}$$

Posons $\mathbf{Dc} = (\alpha, \beta)$ et $\mathbf{Dl} = (\gamma, \delta)$. Notons

$$\mathbf{v}_g = (a_g, b_g) = \mathbf{v}(x_g, y_g).$$

La variable e , introduite en (7), devient alors dans ce contexte :

$$e(x, y) = [\mathbf{v}_g + (x - x_g) \mathbf{Dc} + (y - y_g) \mathbf{Dl} - \mathbf{v}(x, y)] \cdot \nabla f(x, y)$$

On utilise à nouveau la relation (1). On aboutit ainsi à l'expression suivante pour e :

$$(11) \quad e = a_g \cdot \partial f / \partial x + b_g \cdot \partial f / \partial y + \partial f / \partial t + \alpha (x - x_g) \partial f / \partial x + \beta (x - x_g) \partial f / \partial y + \gamma (y - y_g) \partial f / \partial x + \delta (y - y_g) \partial f / \partial y$$

L'expression e dépend maintenant des six paramètres $(a_g, b_g, \alpha, \beta, \gamma, \delta)$, en fonction desquels le critère de décision sera à optimiser. L'explicitation du critère suit le même développement que celui exposé dans le cas du modèle à vitesse constante. La technique d'agglomération des régions est également la même.

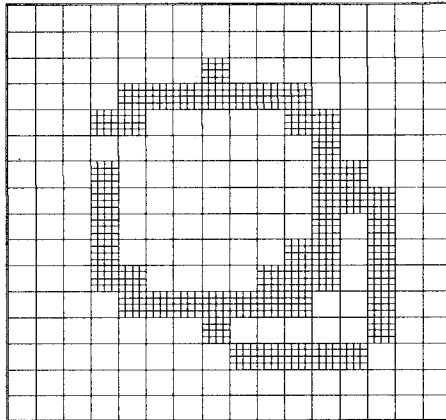
Notons que ce modèle est en fait assez large. Il inclut par exemple le cas d'une combinaison d'une translation et d'une rotation dans le plan image, ainsi que d'une dilatation, cas qui est pris en compte dans [29]. On a alors $\alpha = \delta$ et $\beta = -\gamma$. Par ailleurs, des modèles incluant des aspects 3D (mouvements rigides dans la scène) pourraient aussi être considérés, puisqu'à ce niveau une première segmentation a déjà été réalisée. La technique définie autorise en fait la prise en compte de nombreux modèles selon l'information *a priori* disponible.

Enfin, il pourrait être avancé qu'une segmentation selon un critère uniquement spatial pourrait être considérée, au moins dans une première phase, une étape complémentaire lui étant adjointe selon un critère spatio-temporel. On ne ferait en fait que substituer à la phase initiale de subdivision simple, évoquée précédemment, une procédure beaucoup plus conséquente de segmentation spatiale qui de plus introduirait une sous-partition inutile dans de nombreuses zones. En fait, un tel type de procédure pourra être utilement ajouté à la suite de la segmentation spatio-

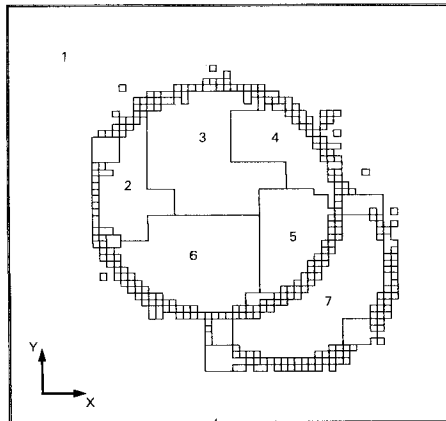
temporelle, ici décrite, afin de gérer la fusion des blocs-frontières résiduels avec les régions obtenues.

4.4. RÉSULTATS

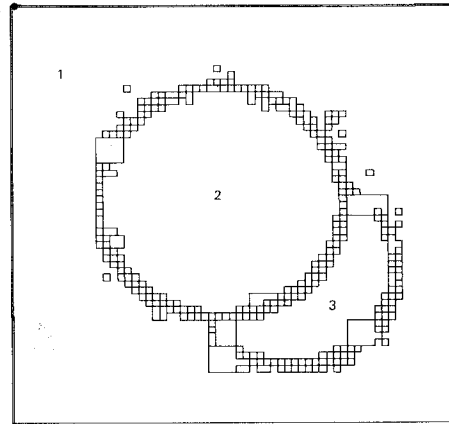
Des détails sur la procédure de type division-et-agglomération utilisée pour réaliser cette segmentation en régions et sur sa mise en œuvre effective, sont donnés dans [30], ainsi que d'autres exemples. Nous allons présenter les résultats de ce procédé sur l'exemple des « deux disques », déjà évoqué dans le cadre de la détermination des ECMs, cf. figure R1 a.



(a)



(b)



(c)

Fig. R4. — (a) Partition à l'issue de la phase 1 : division avec modèle vitesse constante. (b) Partition à l'issue de la phase 2 : agglomération avec modèle vitesse constante. (c) Partition finale à l'issue de la phase 3 : fusion avec modèle plus élaboré.

L'état de la partition est montré à l'issue des trois phases du module, à savoir division initiale en blocs 16×16 , puis éventuellement en blocs 4×4 , figure R4 a, agglomération selon le modèle de vitesse constante, figure R4 b, et enfin agglomération selon le modèle de vitesse plus élaboré, figure R4 c. Le seuil λ_{reg} pour le critère de décision (10) vaut 6 dans les trois phases, ce qui montre bien que l'enrichissement du modèle a bien joué son rôle pour l'agglomération des sous-régions initialement trouvées à l'issue de la phase 2 pour le disque qui se dilate.

Précisons que pour les trois phases, un facteur de normalisation, fonction de la taille de la plus petite des deux zones Z_1 et Z_2 considérées, est introduit dans le seuil. Le tableau de la figure R5 a rassemble les valeurs optimales des paramètres (\hat{a}_0, \hat{b}_0) à l'issue de la phase 2 pour les principales régions du disque se dilatant et pour la région correspondant au disque se translatant. Notons que, même si le propos est de réaliser une segmentation spatio-temporelle et non une identification véritable du mouvement, ces valeurs sont tout à fait cohérentes avec le mouvement entre les deux images pour les régions considérées. Rappe-

Région	a_0	b_0
1	0,0272	0,0188
2	-2,2814	0,7074
3	-0,1421	2,2561
4	2,092	1,215
5	1,936	-1,280
6	-0,812	-2,190
7	2,792	-2,792

(a)

Région	a_g	b_g	α	β	γ	δ
2	0,0823	-0,0426	0,04	-0,0002	-0,0001	0,040

(b)

Fig. R5. — (a) Valeurs optimales des paramètres a_0, b_0 pour la constitution des régions considérées de la figure R4 b. (b) Valeurs optimales des paramètres du modèle plus élaboré lors de la formation de la région 2 de la figure R4 c.

lons que le fond est bruité de façon non corrélée d'une image à l'autre, ce qui explique les faibles écarts obtenus, de l'ordre de 0,02, par rapport à (0, 0), pour la région correspondante. Cette cohérence des résultats peut aussi être relevée pour les valeurs du tableau de la figure R5b qui correspondent aux valeurs optimales à l'issue de la phase 3 pour la région correspondant au disque qui se dilate. (Une dilatation revient en effet à poser $\alpha = \delta$ et $\beta = -\gamma = 0$.) Les blocs 4×4 dits blocs-frontières sont compris dans la phase d'agglomération. Ceux qui restent à l'issue des trois phases pourraient être assimilés aux régions adéquates selon un critère uniquement spatial comme la moyenne, le min-max, et la variance de la distribution des niveaux de gris.

5. Estimation du champ des vitesses le long d'une chaîne contour

Nous supposons que nous disposons d'un ensemble de chaînes contours, fermées ou non. Le propos du module que nous allons décrire est d'estimer, en chaque point de ces chaînes, la deuxième composante du vecteur vitesse, à savoir la composante tangentielle au contour. Ainsi, sera reconstruit le long des contours dans l'image le champ des vitesses apparentes.

Les contraintes fournies par la combinaison, le long des lignes contours, des mesures des paramètres (θ, v^\perp) , issues du module initial « Détermination locale des ECMs », vont permettre d'estimer complètement le champ des vitesses. Cette estimation est effective à condition que des variations suffisantes apparaissent dans les directions spatiales locales des éléments de contour. (Ainsi, une droite contour, sans autre information complémentaire, reste un cas de singularité.)

En fait, le problème est posé en termes de minimisation d'une fonctionnelle simple. La minimisation est mise en œuvre à l'aide d'une méthode de gradient stochastique. Toutefois, il n'est en fait pas garanti *a priori* que le champ apparent observé le long d'une chaîne contour soit cohérent. Aussi, l'estimateur a été complété par un détecteur de Hinkley afin de repérer si des sous-champs de vitesse différents sont présents le long du contour. Ce genre de configuration peut se rencontrer dès qu'une partie du contour résulte en fait d'une occultation, comme dans le cas d'une portion du fond recouvert par un objet en mouvement ou d'un objet passant devant un autre objet. Sur cette partie de la frontière, l'information de mouvement apparent appartient alors à l'objet occultant et non à celui considéré.

5.1. ESTIMATION RÉCURRENTÉ DU CHAMP DES VITESSES

Soit $\omega = (\omega_x, \omega_y)$ le champ des vitesses apparentes dans le plan image à estimer. Considérons l'expression suivante :

$$(12) \quad \varepsilon_s(\omega) = \omega \cdot \mathbf{n}_s - v_s^\perp$$

où

– \mathbf{n}_s représente le vecteur unitaire normal au contour au point d'abscisse s . $\mathbf{n}_s = (-\sin \theta_s, \cos \theta_s)$, l'angle θ_s ayant été dérivé lors du traitement local initial;

– v_s^\perp désigne la composante de la vitesse perpendiculaire au contour initialement mesurée au point s .

On suppose que $\varepsilon_s(\omega)$ est une variable aléatoire stationnaire. Estimer le champ de vitesses ω , le long d'une chaîne de contour C donnée, revient à minimiser la fonctionnelle suivante :

$$(13) \quad J(\omega) = \frac{1}{2} E_C [\varepsilon_s(\omega)^2]$$

où E_C désigne l'espérance considérée sur C .

Cela signifie que pour le champ de vitesses adéquat ω^* , $\varepsilon_s(\omega^*)$ est une variable aléatoire centrée dont la variance est minimale. Des détails sur cette mise en forme du problème peuvent être trouvés dans [21].

Un algorithme de gradient stochastique est alors mis en œuvre afin de réaliser la minimisation de $J(\omega)$. L'estimation s'effectue de façon récurrente de point en point le long de la chaîne. Plus précisément, elle est donnée par :

$$(14) \quad \begin{cases} \omega_0 = v_0^\perp \mathbf{n}_0 \\ \omega_{s+1} = \omega_s - \varepsilon_s(\omega_s) \nabla_{\omega} \varepsilon_s \Gamma \end{cases}$$

où Γ est une matrice de gains et ∇_{ω} note le gradient par rapport à ω . On a :

$$\nabla_{\omega} \varepsilon_s = \left(\frac{\partial \varepsilon_s}{\partial \omega_x}, \frac{\partial \varepsilon_s}{\partial \omega_y} \right) = (-\sin \theta_s, \cos \theta_s)$$

Une argumentation théorique d'une telle formulation de minimisation est montrée dans [32]. Comme le terme correcteur est local, des champs de déplacements variant de point en point peuvent être pris en compte, même si le modèle sous-jacent au critère posé est plutôt du type *constant + bruit*. Plus les coefficients pondérateurs de la matrice Γ sont élevés, meilleure est l'adaptation de l'estimateur; bien sûr un compromis doit être trouvé pour fixer les valeurs de Γ en raison du bruit.

Notons que cet algorithme de minimisation est nettement moins coûteux en calcul que celui retenu dans [23]. Sa convergence est rapide. Le terme traduisant la contrainte de lissage de la vitesse n'est en fait pas explicitement formulé, comme dans [23], mais intervient d'une certaine façon par la récurrence effectuée de point en point le long de la chaîne. Il n'est pas fait recours non plus à des modèles 3D, comme dans [33], nécessitant des hypothèses complémentaires de rigidité et de planarité locale. Enfin, la récurrence définie autorise la mise en œuvre en parallèle d'un détecteur de rupture permettant de mettre en évidence éventuellement des « sous-champs » de vitesse différents le long du contour.

5.1.1. Résultats

Avant de décrire ce détecteur basé sur un test de Hinkley, dont une première présentation est fournie dans [34], nous allons montrer des résultats d'estimation du champ des vitesses le long d'un contour. Le premier exemple illustre le cas d'un déplacement

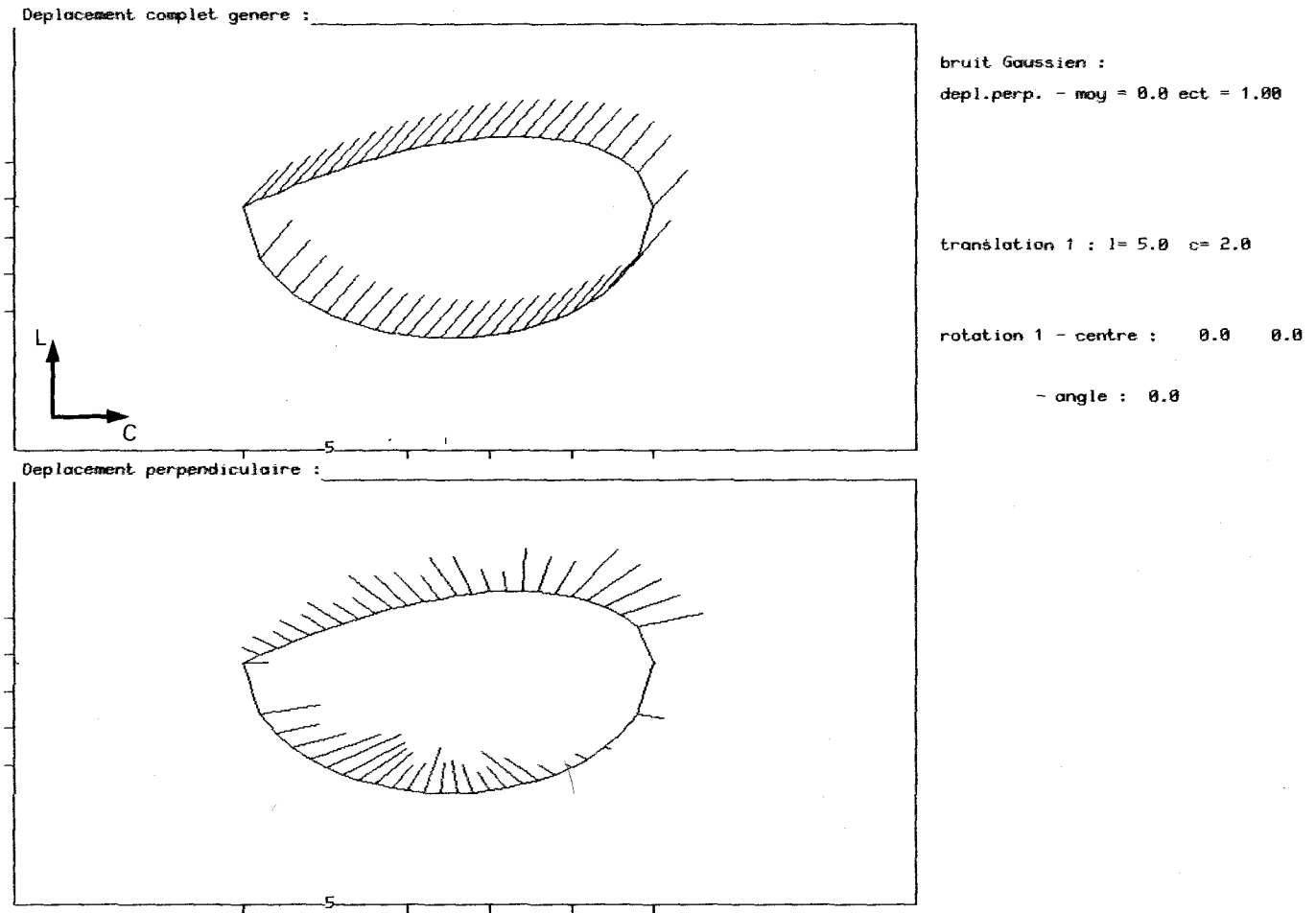


Fig. R6. - Mouvement translationnel pour une courbe, union d'une cubique et d'un morceau d'ellipse.

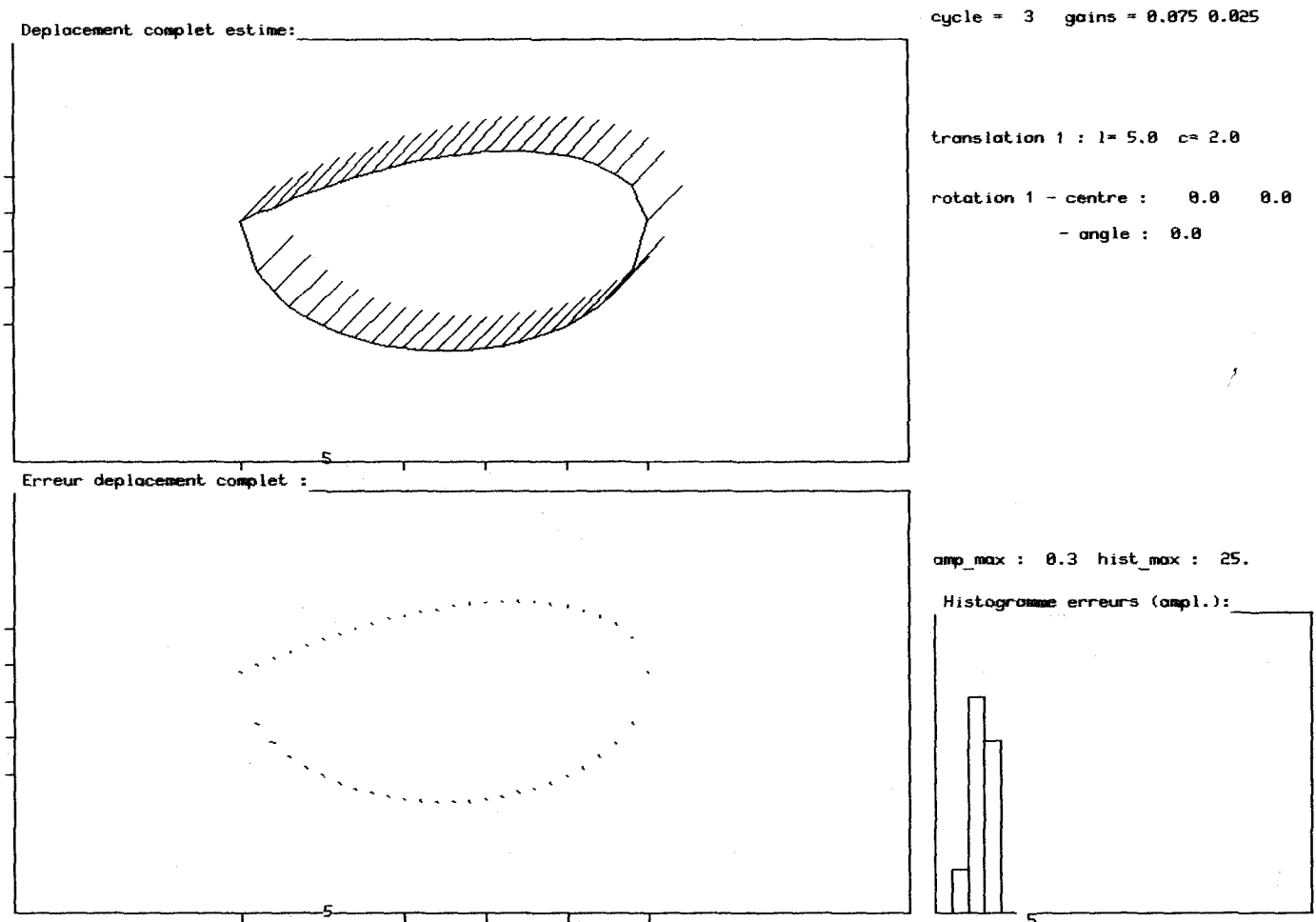


Fig. R7. - Résultats d'estimation du champ des vitesses pour l'exemple de la figure R6.

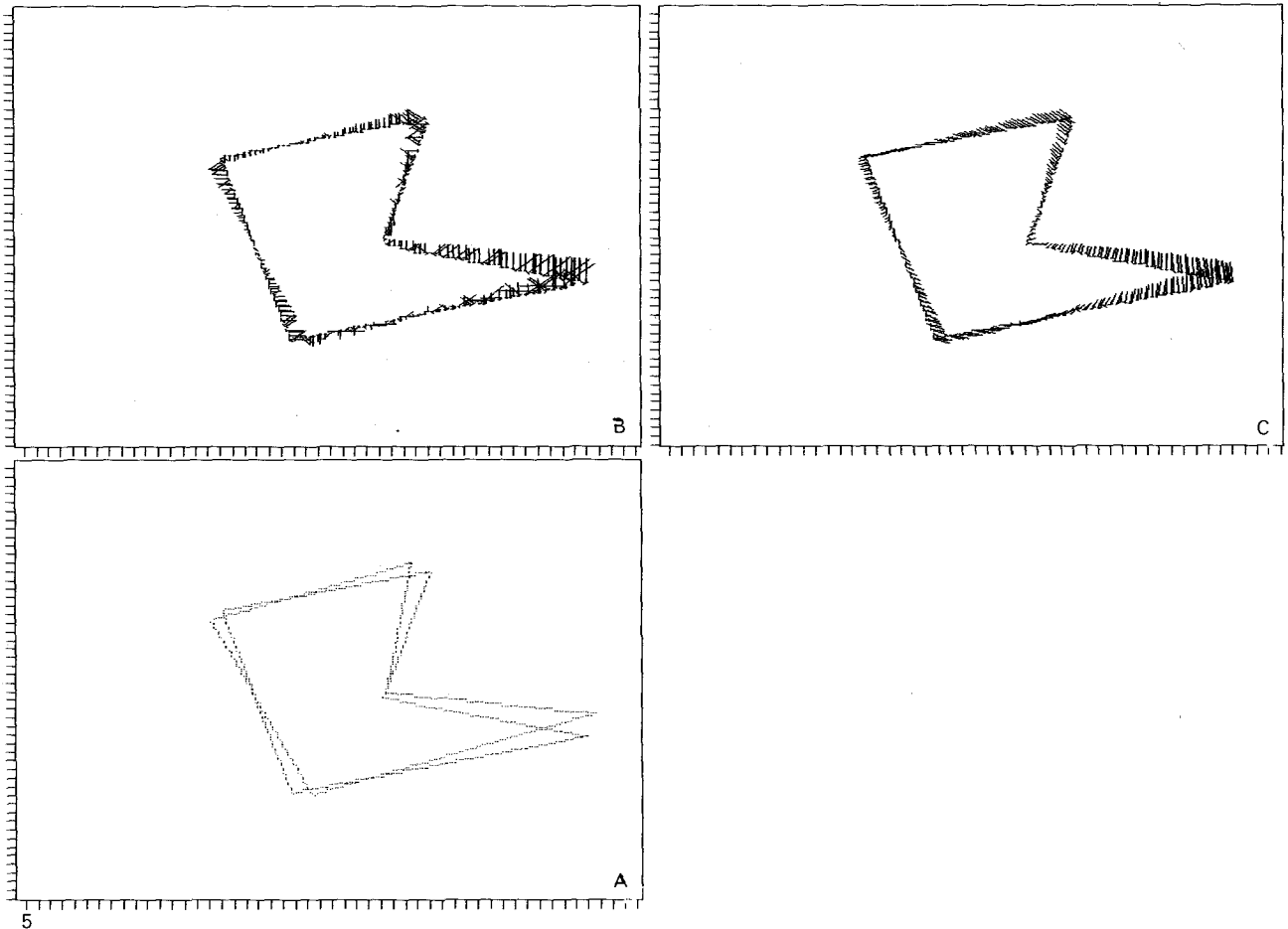


Fig. R8. — Cas d'un mouvement de rotation. (a) Silhouettes surimposées du polygone dans les deux images successives.

(b) ECMs déterminés. (c) Champ des vitesses estimé, $\Gamma = \begin{bmatrix} 0,04 & 0,01 \\ 0,01 & 0,04 \end{bmatrix}$, deux cycles.

translational. Il a été généré le long d'une courbe fermée, constituée de l'union d'une cubique et d'un morceau d'ellipse, et échantillonnée en 50 points. Le champ des déplacements perpendiculaires est bruité selon un bruit gaussien, figure R6. (Pour une bonne compréhension de cette figure ainsi que d'un certain nombre de figures suivantes, précisons que les échelles sur les deux axes ne sont pas nécessairement les mêmes.) Les résultats d'estimation ainsi que les erreurs entre champ exact et champ estimé sont donnés à la figure R7. Les deux valeurs de gains fournies correspondent, la première, aux deux éléments diagonaux de la matrice Γ , la seconde aux deux éléments transverses. La chaîne étant fermée, la récurrence s'effectue de point en point de manière cyclique. Dans le cas contraire, on procéderait par aller-retour. Dans l'exemple considéré, trois cycles ont été réalisés.

Le deuxième exemple traite le cas d'un mouvement de rotation et illustre ainsi le pouvoir de « poursuite » du gradient stochastique et de lissage du bruit, figure R8. Il s'agit de deux images synthétiques d'un polygone. La superposition des deux silhouettes du polygone illustre simplement le déplacement produit entre les deux images. Dans cet exemple, le champ des vitesses perpendiculaires le long de la frontière du polygone est obtenu par la méthode exposée dans ce papier, en considérant les deux images intensités

successives. L'aspect quelque peu chaotique du champ perpendiculaire déterminé résulte notamment d'effets de « marches d'escalier » dans les segments de droite de la frontière. Deux cycles d'itérations ont été menés dans les deux sens de parcours de la frontière de l'objet, sens trigonométrique et sens inverse, à l'issue desquels est réalisé un moyennage des deux estimations obtenues.

5. 2. DÉTECTION DES LIEUX DE DISCONTINUITÉ

Le test de Hinkley est un test à « somme cumulée » qui permet de détecter des sauts de moyenne du processus observé [35]. Ici, ce dernier est défini par :

$$(15) \quad y_s = \mathbf{v}_{s-1} \cdot \mathbf{n}_s - v_s^\perp$$

où s désigne l'abscisse curviligne du point courant sur la courbe contour considérée, \mathbf{n}_s le vecteur normal unitaire en ce point et v_s^\perp la composante perpendiculaire de la vitesse issue de l'estimateur local d'ECM, \mathbf{v}_{s-1} le vecteur vitesse estimé au point précédent $s-1$. Le principe du détecteur de Hinkley est le suivant. Soit μ_0 la moyenne du processus observé avant la rupture. En fait, deux détecteurs sont activés en parallèle. Le premier permet de surveiller les sauts de moyenne vers le bas, c'est-à-dire de la valeur μ_0 vers

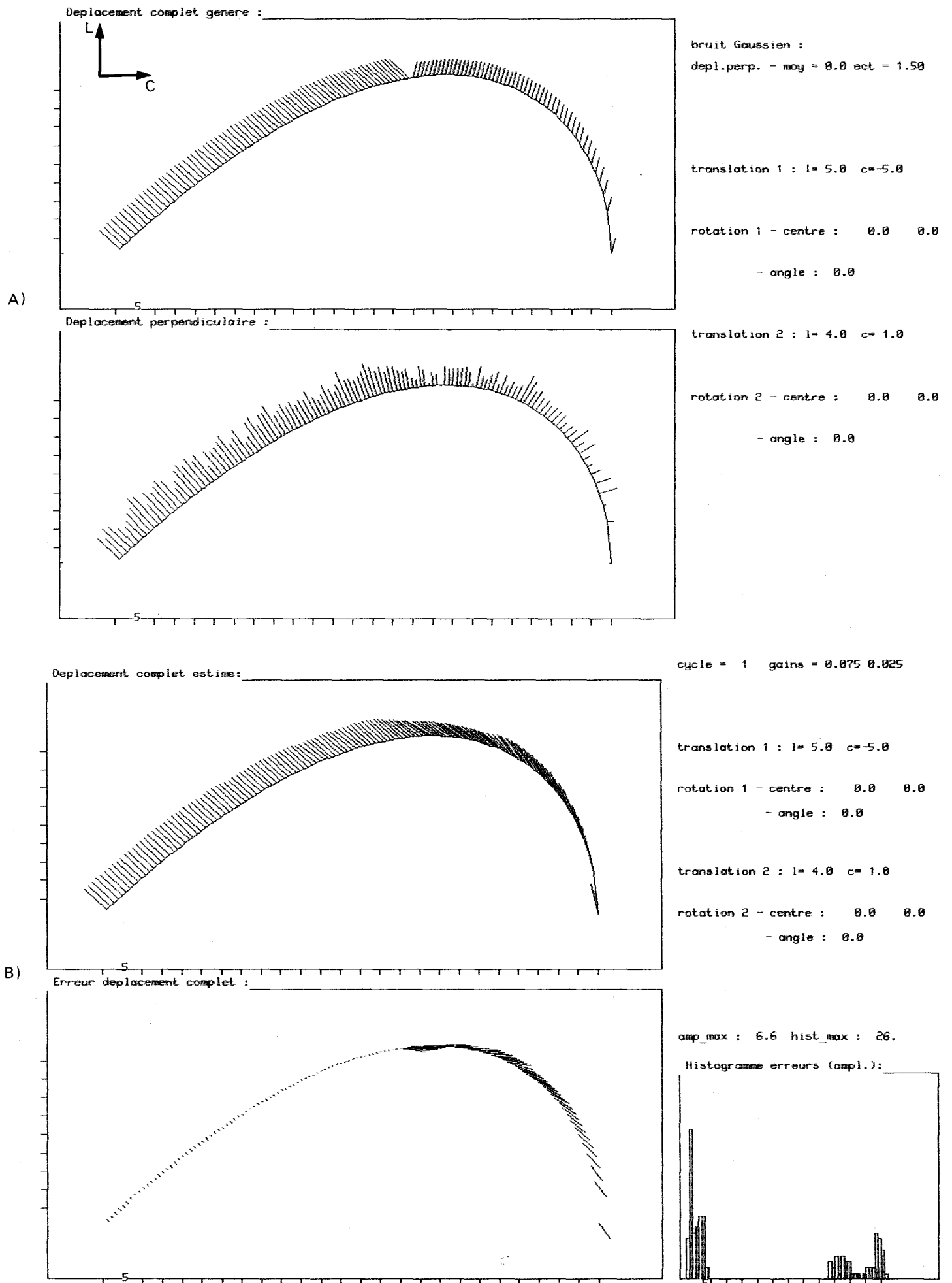


Fig. R9. - (a) Cas de sous-champs de vitesse differents.
(b) Estimation sans test de Hinkley.

une valeur μ_1 , telle que $\mu_1 < \mu_0 - \Delta\mu$. Il est donné par les formules suivantes :

$$(16 a) \quad R_s = \sum_{i=1}^s (y_i - \mu_0 + \Delta\mu)$$

$$(16 b) \quad P_s = \max_{0 \leq k \leq s} R_k \quad (R_0 = 0)$$

La détection intervient lorsque $|P_s - R_s| > \lambda_d$ où λ_d est un seuil prédéterminé, de même que $\Delta\mu$. La position estimée \tilde{s} du saut correspond alors au dernier point où $|P_s - R_s| = 0$.

Une définition similaire peut être donnée pour le second détecteur. Ce dernier surveille des sauts de moyenne vers le haut, c'est-à-dire de la valeur μ_0 vers une valeur μ_2 , telle que $\mu_2 > \mu_0 + \Delta\mu$. Son expression est comme suit :

$$(17 a) \quad r_s = \sum_{i=1}^s (y_i - \mu_0 - \Delta\mu)$$

$$(17 b) \quad p_s = \min_{0 \leq k \leq s} r_k \quad (r_0 = 0)$$

Le critère de détection est semblable, à savoir $|p_s - r_s| > \lambda_d$. La même définition vaut pour la localisation de l'endroit de rupture.

Appliqué à notre problème, ce test s'interprète de la façon suivante. Tant que l'estimation par récurrence

concerne un champ de vitesses cohérent, la moyenne de y est proche de 0. Dans le cas contraire, elle diverge. Un saut est évidemment détecté avec un certain retard. La récurrence de l'estimation, selon l'algorithme de gradient stochastique, est alors réinitialisée au point de rupture trouvé plus un. Les détecteurs sont également réinitialisés.

5.2.1. Résultats

Les figures R9 à R11 illustrent cette procédure de détection de rupture. Deux champs de translation différents sont produits le long d'une cubique. Le champ perpendiculaire résultant est bruité avec un bruit Gaussien, figure R9a. Si, à partir de ce champ, on applique l'estimateur récurent tel quel, une convergence rapide se produit tant que l'on reste dans la portion correspondant au premier type de déplacement, puis les estimations deviennent nettement erronées dès que la seconde portion est abordée, figure R9b. On a donc activé un détecteur de Hinkley avec les paramètres suivants, $\Delta\mu = 1,5$ et $\lambda_d = 5$. La moyenne μ_0 est posée égale à 0. (Notons que d'autres expérimentations ont été conduites avec une moyenne μ_0 récursivement estimée.) De plus, le processus observé est filtré afin d'éviter des fausses alarmes dues en fait à des erreurs importantes isolées. Le filtre suivant est considéré : $y_s = \text{sign}(y_s) \min(|y_s|, \delta)$, avec $\delta = 5$. La figure R10 montre le comportement du test, la détection s'effectuant bien au point $s = 74$. Enfin,

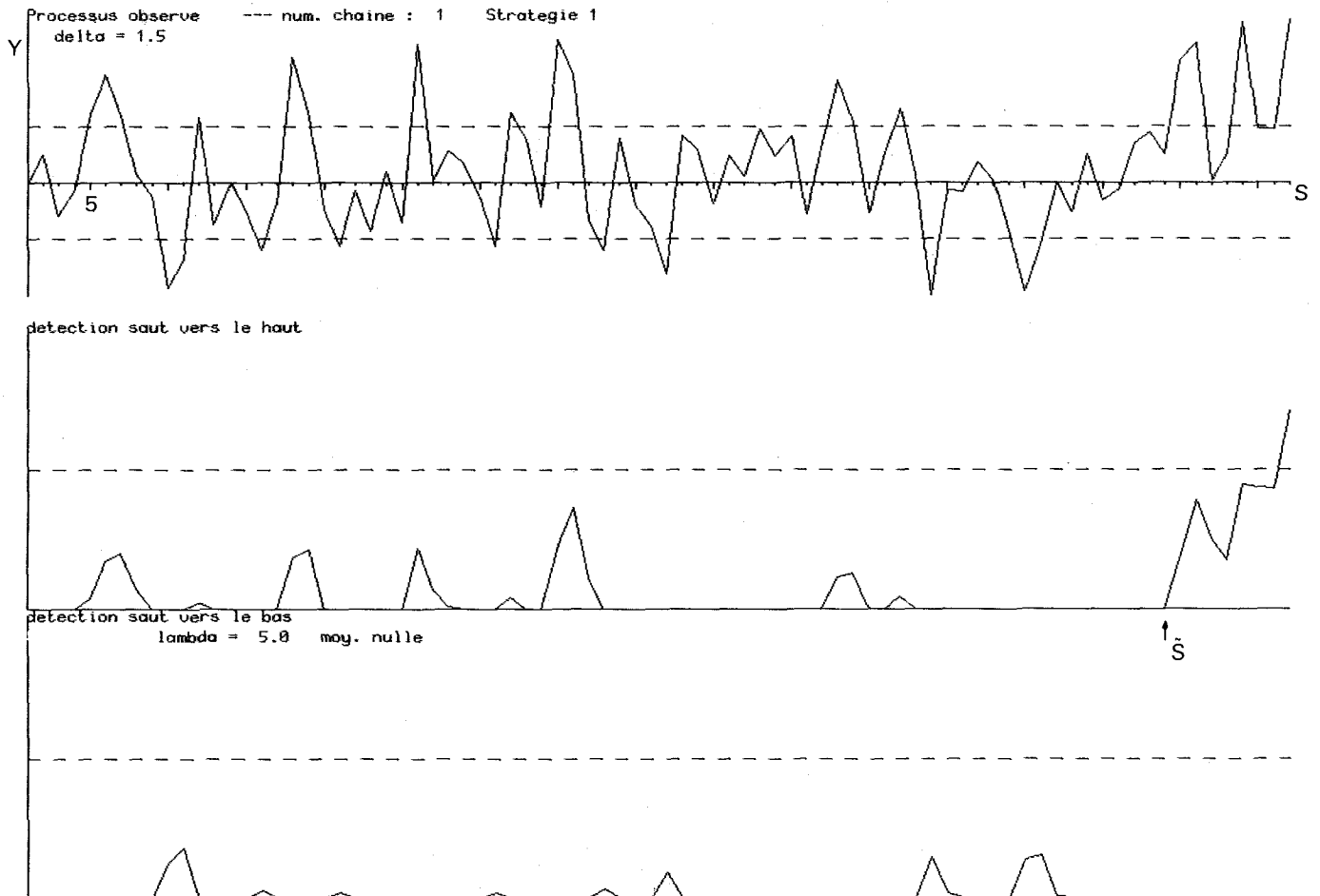


Fig. R10. - Comportement du détecteur de Hinkley pour l'exemple de la figure R9a.

cycle = 1 gains = 0.075 0.025

translation 1 : l= 5.0 c=-5.0

rotation 1 - centre : 0.0 0.0
- angle : 0.0

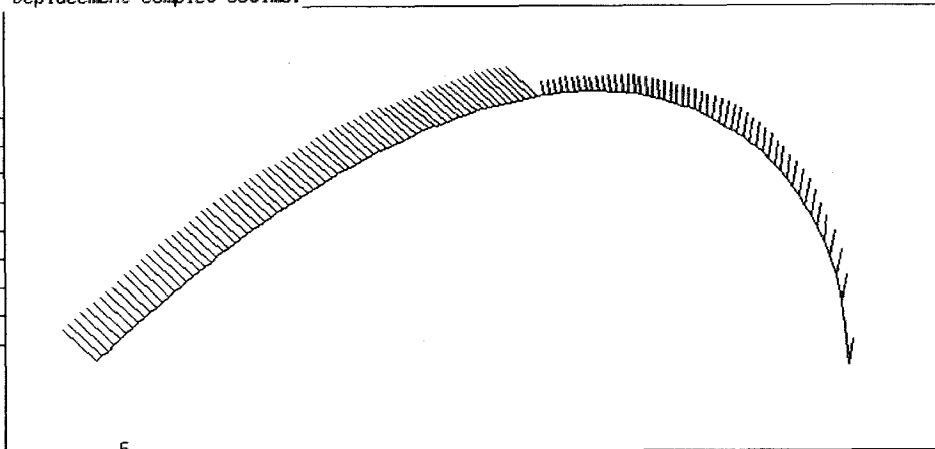
translation 2 : l= 4.0 c= 1.0

rotation 2 - centre : 0.0 0.0
- angle : 0.0

amp_max : 2.1 hist_max : 26.

Histogramme erreurs (ampl.):

Deplacement complet estime:



Erreur deplacement complet :

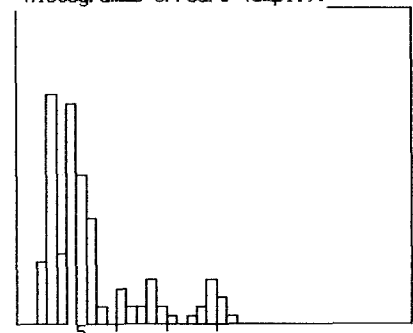
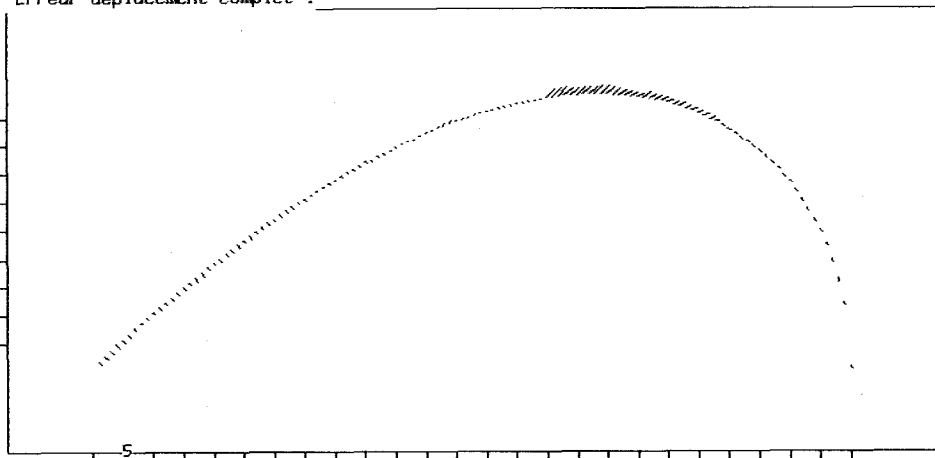


Fig. R11. — Estimation du champ des vitesses avec test de Hinkley et réinitialisation.

est tracé à la figure R11 le champ des vitesses maintenant obtenu en prenant en compte le test précédent et en réinitialisant l'estimation récurrente au point de discontinuité détecté (après un seul cycle).

6. Compléments et conclusion

6.1. MODULE FUSION

Nous allons succinctement apporter quelques compléments concernant les modules du schéma d'ensemble, décrit à la figure 1, qui n'ont pas été abordés dans le corps de cet article.

Le module *Fusion* reste à réaliser dans le cas où les approches contour et région sont conjointement mises en œuvre. A l'issue des deux procédés de détermination locale des ECMs et l'obtention des régions spatio-temporelles, les deux lots suivants de structures sont obtenus, des chaînes de contour (avec le champ de vitesses normales associé) d'une part, des zones étiquetées (selon un critère de mouvement apparent) d'autre part. Précisons que le chaînage s'effectue uniquement selon les caractéristiques spatiales des ECMs; aussi un algorithme de chaînage de contours spatiaux convient.

Il s'agira d'établir un processus coopérant permettant de corriger éventuellement ces primitives (citons par

exemple le cas d'une composante perpendiculaire de la vitesse trouvée quasi isolément non nulle au sein d'une région déclarée fixe), de fermer les chaînes contours, d'attribuer les blocs frontières élémentaires restants aux régions appropriées, de trier entre contours intérieurs et contours frontières, ... L'objectif est aussi d'aboutir à une structure de données représentant la partition de l'image, riche, efficace, facilement manipulable pour fournir un important lot d'informations aussi bien immédiates que résultantes.

6.2. MODULE ESTIMATION DU CHAMP DES VITESSES SUR UN DOMAINE

Comme estimateur du champ des vitesses au sein d'un domaine borné est repris celui présenté dans [36]. Il s'agit d'un schéma d'estimation à deux indices utilisant la notion des modèles complémentaires, s'appuyant sur une modélisation stochastique du champ de vitesses, et prenant en compte les valeurs aux frontières du domaine. Une variance est associée à l'estimé du vecteur vitesse en chaque point de la frontière. Deux cas typiques sont à noter. Une variance nulle correspond à un vecteur vitesse considéré comme exact (condition de Dirichlet), une variance infinie à une absence d'information (condition de Neumann). Ce dernier aspect est d'importance car il correspond aux portions occultées de la frontière

d'une région. La mise en œuvre de l'estimation, nécessitant la résolution d'équations de Poisson couplées, qui sont d'ailleurs les mêmes que celles auxquelles aboutit la méthode de régularisation développée dans [17], est basée sur une technique de relaxation locale.

6. 3. CONCLUSION

Nous avons exposé une méthodologie pour l'extraction du mouvement apparent dans une séquence d'images numériques, problème dont l'un des aspects essentiels est la gestion des discontinuités potentielles. Elle intègre à la fois des aspects de segmentation et d'estimation et fournit un ensemble d'indices spatio-temporels de différents niveaux, tout en conduisant à l'estimation du champ des vitesses sur toute l'image et à l'obtention de zones délimitées et homogènes au sens du mouvement.

La détermination locale d'éléments de contour en mouvement s'appuie sur une modélisation d'une portion de surface dans l'espace (x, y, t) et un schéma fondé sur un test d'hypothèses, dont la complexité de mise en œuvre effective est équivalente à un simple détecteur de contour spatial de type gradient-convolution. Les contours dits d'occlusion entrent de la même façon dans ce schéma et l'amplitude de déplacement mesurable n'est pas bornée *a priori*. Parallèlement, une procédure de segmentation en régions spatio-temporelles a été définie à partir d'un critère ne comprenant qu'une information partielle de mouvement. Cette procédure est hiérarchique au sens où une succession de modèles de plus en plus élaborés peut intervenir, et ce à travers le même formalisme basé sur un test de vraisemblance. Enfin, nous avons proposé une méthode récurrente utilisant un algorithme de gradient stochastique pour l'estimation complète du champ des vitesses le long des chaînes contours, et autorisant en parallèle la prise en compte d'un test de Hinkley pour la détection de sous-champs différents éventuels. L'ensemble de ces modules a été validé sur de nombreux exemples comprenant des données synthétiques bruitées et des images réelles.

Remerciements

La première partie de cette étude entre dans le cadre d'une convention ADI, n° 84/805 projet 1707. Par ailleurs, l'auteur tient à remercier Albert Benveniste pour de nombreuses discussions et Javier Santillana Rivero pour sa collaboration sur le problème de la segmentation en régions. Les lecteurs sont également remerciés pour leurs commentaires. Enfin, notons que l'étude référencée en [36] résulte de contacts entre le LIDS du MIT et l'IRISA.

Manuscrit reçu le 23 décembre 1986.

BIBLIOGRAPHIE

- [1] C. CAFFORIO et F. ROCCA, Methods for measuring small displacements of television images, *IEEE-Trans. Information Theory*, IT-22, n° 5, sept. 1976, p. 573-579.
- [2] A. N. NETRAVALI et J. D. ROBBINS, Motion-compensated television coding: Part I, *Bell Syst. Tech. Jnl*, 58, n° 3, mars 1979, p. 631-670.
- [3] B. CHOQUET et J. SANTILLANA RIVERO, Algorithmes de détection et d'estimation de mouvement des images de télévision, *RT-CCETT/RTI*, 007-85, mars 1985.
- [4] J. A. LEESE, C. S. NOVAK et B. B. CLARK, An automated technique for obtaining cloud motion from geosynchronous satellite data using cross-correlation, *Jal of Applied Meteorology*, 10, fév. 1971, p. 118-132.
- [5] A. L. GILBERT, M. K. GILES, G. M. FLACHS, R. B. ROGERS et Y. HSUN, A real-time video tracking system, *IEEE-Trans. Pattern Analysis and Machine Intelligence*, PAMI-2, n° 1, janv. 1980, p. 47-56.
- [6] T. S. HUANG éd., Image Sequence Processing and Dynamic Scene Analysis, *NATO-ASI Series F2*, Springer-Verlag, 1983.
- [7] J. BARON, *A survey of approaches for determining optical flow, environmental layout and egomotion*, Univ. of Toronto, Dept. of Computer Science, RBCVTR-84-5, nov. 1984.
- [8] R. JAIN, *Dynamic scene analysis*, Univ. of Michigan, Computer Research Lab., CRL-TR-6-84, janv. 1984, aussi in *Progress in Pattern Recognition 2*, L. KANAL et A. ROSENFELD éd., North-Holland, 1986.
- [9] H.-H. NAGEL, Image sequences—Ten (octa-)years—From phenomenology towards a theoretical foundation, *Proc. 8^e Int. Conf. on Pattern Recognition*, Paris, oct. 1986, p. 1174-1185.
- [10] P. RIVES, E. BREUIL et B. ESPIAU, Recursive estimation of 3D features using optical flow and camera motion, *Proc. Conf. Intelligent Autonomous Systems*, Amsterdam, déc. 1986.
- [11] J. TSOTSOS, Temporal event recognition: an application to left ventricular performance assessment, *Proc. Int. Joint Conf. on Artificial Intelligence*, Vancouver, IJCAI-1981, p. 900-907.
- [12] A. WAXMAN et S. ULLMAN, Surface structure and three-dimensional motion from image flow kinematics, *The Int. Jnl of Robotics Research*, 4, n° 3, 1985, p. 72-94.
- [13] K. PRAZDNY, On the information in optical flows, *Computer Vision, Graphics & Image Processing*, CVGIP-22, 1983, p. 239-259.
- [14] A. MITICHE, Computation of optical flow and rigid motion, *Proc. 2nd IEEE Workshop on Computer Vision: Representation and Control*, Annapolis, 1984, p. 63-71.
- [15] G. ADIV, Determining three-dimensional motion and structure from optical flow generated by several moving objects, *IEEE-Trans. on Pattern Analysis and Machine Intelligence*, PAMI-7, n° 4, juil. 1985, p. 384-401.
- [16] S. T. BARNARD et W. B. THOMPSON, Disparity analysis of images, *IEEE-Trans. Pattern Analysis and Machine Intelligence*, PAMI-2, n° 4, juil. 1980, p. 333-340.
- [17] B. K. P. HORN et B. G. SCHUNCK, Determining optical flow, *Artificial Intelligence*, 17, 1981, p. 185-203.
- [18] P. ANANDAN, Computing dense displacement fields with confidence measures in scenes containing occlusion, *SPIE*, 521, *Intelligent Robots and Computer Vision*, Cambridge, Nov. 1984, p. 184-194.
- [19] B. G. SCHUNCK, Motion segmentation and estimation by constraint line clustering, *Proc. 2nd IEEE Workshop on Computer Vision: Representation and Control*, Annapolis, 1984, p. 58-62.

- [20] H.-H. NAGEL et W. ENKELMANN, An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences, *IEEE-Trans. Pattern Analysis and Machine Intelligence*, PAMI-8, n° 5, sept. 1986, p. 565-593.
- [21] P. BOUTHEMY, Un nouveau schéma d'estimation du champ des vitesses sur les contours dans une séquence d'images, 2^e Colloque Image CESTA, Nice, avril 1986, p. 237-245.
- [22] B. G. SCHUNCK, The image flow constraint equation, *Computer Vision, Graphics & Image Processing*, CVGIP-35, 1986, p. 20-46.
- [23] E. C. HILDRETH, Computations underlying the measurement of visual motion, *Artificial Intelligence*, 23, 1984, p. 309-354.
- [24] B. F. BUXTON et H. BUXTON, Computation of optic-flow from the motion of edge features in image sequences, *Image and Vision Computing*, 2, n° 2, mai 1984, p. 59-75.
- [25] P. BOUTHEMY, Estimation of edge motion based on local modeling, SPIE, 595, *Computer Vision for Robots*, Cannes, déc. 1985, p. 162-169.
- [26] R. JAIN, W. N. MARTIN et J. K. AGGARWAL, Segmentation through the detection of changes due to motion, *Computer Graphics and Image Processing*, CGIP-11, 1979, p. 13-34.
- [27] Y. Z. HSU, H.-H. NAGEL et G. REKERS, New likelihood test methods for change detection in image sequences, *Computer Vision, Graphics Image Processing*, CVGIP-26, 1984, p. 73-106.
- [28] W. B. THOMPSON, K. M. MUTCH et V. A. BERZINS, Dynamic occlusion analysis in optical flow fields, *IEEE-Trans. Pattern Analysis and Machine Intelligence*, PAMI-7, n° 4, juil. 1985, p. 374-383.
- [29] R. HARTLEY, Segmentation of optical flow fields by pyramid linking, *Pattern Recognition Letters*, 3, juil. 1985, p. 253-262.
- [30] J. SANTILLANA RIVERO et P. BOUTHEMY, Region segmentation according to motion-based criteria, Rapport de Recherche INRIA, n° 629, fév. 1987.
- [31] J. L. HOROWITZ et T. PAVLIDIS, Picture segmentation by a direct split-and-merge procedure, *Proc. 2nd Int. Conf. on Pattern Recognition*, 1974, p. 424-433.
- [32] A. BENVENISTE, M. GOURSAT et G. RUGET, Analysis of stochastic approximation schemes with discontinuous and dependent forcing terms with applications to data communication algorithms, *IEEE-Trans. Automatic Control*, AC-25, 25, n° 6, déc. 1980, p. 1042-1058.
- [33] G. TZIRITAS, Estimation du flux optique basée sur le gradient, *Traitement du Signal*, 3, n° 1, 1986, p. 3-11.
- [34] P. BOUTHEMY, A method of integrating motion information along contours including segmentation, *Proc. 8^e Int. Conf. on Pattern Recognition*, Paris, oct. 1986, p. 651-653.
- [35] M. BASSEVILLE, On-line detection of jumps in mean, in *Detection of Abrupt Changes in Signals and Dynamical Systems*, M. BASSEVILLE et A. BENVENISTE éd., *Lecture Notes in Control and Information Science*, 77, Springer-Verlag, 1986.
- [36] A. ROUGÉE, B. LEVY et A. S. WILLSKY, *Optic flow estimation inside a bounded domain*, MIT, Lab. for Information and Decision Systems, Rapport LIDS-P-1589, août 1986.