

Estimation du flot optique en présence de discontinuités : une revue

Optical Flow Estimation Preserving Discontinuities : a Survey

par Maciej ORKISZ et Patrick CLARYSSE

CREATIS UMR CNRS (# 5515), affilié à l'INSERM
INSA 502 F-69621 Villeurbanne cedex

résumé et mots clés

L'estimation du mouvement à partir de séquences d'images bidimensionnelles s'appuie sur deux hypothèses de base : l'hypothèse de conservation de la luminance des objets au cours de leurs mouvements et l'hypothèse de continuité spatiale, temporelle ou spatio-temporelle, du champ de vitesses apparentes. Cette dernière hypothèse est valable localement, « à l'intérieur » des objets, mais elle provoque un lissage indésirable au voisinage des frontières entre les projections, dans le plan image, des objets animés de mouvements différents. Ces frontières sont appelées discontinuités de mouvement. Le sujet principal de cet article est la revue des techniques visant à estimer le champ de vitesses apparentes en évitant de lisser les discontinuités de mouvement. La première partie de l'article présente les méthodes s'appuyant sur l'hypothèse selon laquelle les discontinuités de mouvement coïncident spatialement avec certaines frontières photométriques. La deuxième partie rapporte les méthodes de segmentation du champ estimé courant en régions homogènes au sens du mouvement. La troisième partie s'intéresse aux méthodes qui réalisent simultanément l'estimation du champ de vitesses et la détection des ruptures locales de continuité de ce champ. Ceci permet d'inhiber localement le lissage, dans les zones susceptibles de contenir une discontinuité de mouvement. Cette inhibition peut être obtenue par introduction d'un « processus de ligne » binaire, grâce à l'utilisation d'un « estimateur robuste », ou dans un schéma de diffusion anisotrope. La dernière partie est consacrée à la gestion des occultations. Les estimations obtenues dans les zones d'occultation sont erronées à cause de la violation des deux hypothèses de base : conservation des propriétés photométriques et continuité.

Estimation du mouvement, Flot optique, Discontinuités.

abstract and key words

Motion estimation from image sequences is based on two assumptions : the brightness conservation assumption and the assumption of spatial, temporal or spatio-temporal continuity (i.e. smoothness constraint) of the apparent velocity field. The latter assumption holds locally, within the objects, but it results in blurring the boundaries between the projections, onto the image plane, of objects undergoing different motions. These boundaries are called motion discontinuities. The main subject of this paper is an overview of the existing techniques designed to estimate the apparent velocity fields while preserving the motion discontinuities. The first part deals with the methods based on an assumption which states that the motion discontinuities spatially coincide with image brightness boundaries. The second part reports the motion segmentation methods. The third part describes the methods which perform the motion field estimation while detecting the local discontinuities of the currently estimated field so as to avoid smoothing in the areas likely to contain motion boundaries. These discontinuities are preserved by means of a line process, a robust estimator or within an anisotropic diffusion scheme. The last part is devoted to the occlusions. The estimation errors in the occlusion areas are due to the violation of two basic assumptions : the continuity assumption and particularly the conservation assumption.

Motion Estimation, Optical Flow, Discontinuities.

1. introduction

Le mouvement, dans une séquence d'images bidimensionnelles (2D), est perceptible grâce aux changements de la distribution spatiale des intensités lumineuses. Le mouvement ainsi perçu est appelé « *apparent* », car il ne correspond pas nécessairement à la projection, dans le plan de l'image, du mouvement ayant lieu dans l'espace tridimensionnel (3D). Ainsi, par exemple, les vitesses apparentes des points situés sur une sphère uniforme en rotation sont nulles, donc différentes des projections des vraies vitesses de ces points. Le champ des vitesses apparentes porte le nom de « *flot optique* » [Horn et Schunck'81]. Nous utiliserons indifféremment les termes « *flot optique* » et « *champ de vitesses* » pour désigner le champ des vitesses apparentes, bien que certains auteurs nuancent la signification de ces termes en désignant par « *champ de vitesses apparentes* » une approximation du flot optique calculée à l'aide d'un algorithme donné [Mitiche et Bouthémy'96]. Notons que l'article cité ici est une excellente étude de l'état de l'art et des difficultés dans le domaine de l'estimation du mouvement.

L'estimation du mouvement apparent dans une séquence d'images s'appuie nécessairement sur une hypothèse de conservation de certaines propriétés photométriques des objets filmés. Les seules propriétés que l'on peut attribuer à un point pris isolément sont l'intensité lumineuse et, éventuellement, la couleur. De nombreux auteurs ont démontré que l'estimation du mouvement utilisant pour seule contrainte la conservation de l'intensité est un problème sous-déterminé. Sous les hypothèses de « *petits déplacements* » et de différentiabilité spatio-temporelle de l'intensité lumineuse, la contrainte de conservation de l'intensité est généralement exprimée par l'équation de contrainte de mouvement au premier ordre [Horn et Schunck'81] :

$$I_x u + I_y v + I_t = \nabla I \cdot \mathbf{v} + I_t = 0 \quad (1)$$

où I_x, I_y, I_t, u et v représentent respectivement : les composantes horizontale et verticale du gradient spatial de l'intensité ∇I , le gradient temporel de l'intensité et les composantes horizontale et verticale du vecteur vitesse \mathbf{v} . L'équation (1) seule permet de déterminer uniquement la projection du vecteur vitesse dans la direction du gradient spatial de l'intensité. Cette projection étant localement perpendiculaire aux frontières photométriques, on l'appelle « *composante normale* » du vecteur vitesse. Pour trouver la deuxième composante, tangentielle, il est nécessaire de régulariser l'estimation, c'est-à-dire réduire l'espace des solutions par l'introduction d'une contrainte supplémentaire. L'utilisation de la couleur, consistant à écrire l'équation (1) pour chaque composante de la couleur séparément, permet parfois de contraindre suffisamment la solution [Lai et al.'94]. Cependant les trois composantes de la couleur sont souvent corrélées entre elles et n'apportent alors aucune contrainte supplémentaire. Une approche semblable consiste à imposer la conservation de plusieurs grandeurs (contraste, dérivées, moments, courbure, entropie...), ce qui conduit à construire un système multi-contraintes

en écrivant une équation de type (1) pour chacune de ces grandeurs séparément [Wohn et al.'83], [Mitiche et al.'87], [Brofferio et Mercalli'89], [Tistarelli'95]. Plus souvent cependant, on introduit une *contrainte de continuité* (lissage) du champ de vitesses, selon laquelle les points voisins sont animés de mouvements « *très semblables* ». La régularisation conduit, dans ce dernier cas, à la minimisation d'une fonctionnelle comportant un terme d'attache aux données U_d et un terme de lissage U_s , pondérés par un coefficient de régularisation α :

$$U(u, v) = U_d(u, v) + \alpha U_s(u, v) \quad \alpha > 0 \quad (2)$$

avec, par exemple :

$$\begin{aligned} U_d(u, v) &= \iint (I_x u + I_y v + I_t)^2 dx dy \\ U_s(u, v) &= \iint (u_x^2 + u_y^2 + v_x^2 + v_y^2) dx dy \end{aligned} \quad (3)$$

où u_x, u_y, v_x, v_y représentent les gradients horizontaux et verticaux de la vitesse apparente [Horn et Schunck'81]. Différentes formulations de contraintes de conservation et de lissage, alternatives par rapport à (1) et (2), sont passées en revue dans [Singh'91].

Dans les deux approches, avec un système multi-contraintes et avec une contrainte de lissage, on fait une hypothèse sous-jacente selon laquelle l'ensemble des points utilisés pour lisser le champ ou pour déterminer la grandeur à conserver, appartient à un même objet. Cette hypothèse n'est pas respectée en présence de *discontinuités de mouvement* (fig. 1). Par discontinuité de mouvement, dans les images 2D, on entend une frontière entre les projections d'objets différents animés de mouvements indépendants ou entre les projections de différentes parties d'un même objet articulé ou non rigide (cas du geste humain, par exemple).

L'objet principal de cet article est la revue des techniques qui visent à estimer le champ de vitesses apparentes tout en préservant les discontinuités de mouvement. La littérature traitant de ce sujet est abondante et la bibliographie proposée ne peut

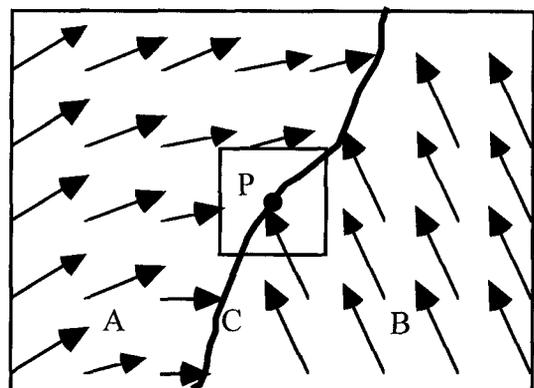


Figure 1. – Exemple de discontinuité de mouvement : tout voisinage du point P situé sur la frontière C contient des points appartenant aux deux objets A et B animés de mouvements différents.

être exhaustive. Elle s'appuie cependant sur une centaine de références sélectionnées, qui permettent de dégager les approches les plus marquantes des quinze dernières années, insistant sur celles qui semblent aujourd'hui les plus « solides ». Quelques méthodes récentes, moins connues, mais originales et prometteuses, sont également mentionnées. Le caractère de l'article est essentiellement tutoriel, sans ambition de comparaison de performances des méthodes citées. En effet, il est difficilement envisageable de dresser une comparaison exhaustive et cohérente d'un grand nombre de méthodes fondées sur des principes variés. D'ailleurs, il n'existe pas, à notre connaissance, de tentative semblable dans la littérature, excepté [Barron *et al.*'94] qui concerne des méthodes ne prenant pas en compte les discontinuités.

Outre l'introduction et les conclusions, l'article est organisé en quatre parties dont les trois premières sont plus particulièrement consacrées aux différentes techniques visant à prendre en compte la présence des discontinuités de mouvement. Une localisation parfaite des discontinuités de mouvement devrait permettre de délimiter la projection de chaque objet animé d'un mouvement distinct. Ainsi serait-il possible de prendre en compte tous les points appartenant à un objet et seulement ces points, afin de déterminer le mouvement apparent de cet objet. En pratique cependant, non seulement la segmentation des images en régions homogènes au sens du mouvement est inconnue a priori, mais au contraire un des objectifs de l'estimation du mouvement est souvent d'effectuer une telle segmentation. *L'estimation et la segmentation du mouvement sont donc interdépendantes* [Bouhémy'95]. On peut cependant distinguer une approche mettant l'accent sur la segmentation et une approche mettant au premier plan l'estimation. La première approche prend en compte les discontinuités par essence même, mais de façon *globale*, comme lieu de frontières entre les régions de la segmentation. La deuxième approche manipule des champs denses conjointement à un marquage, binaire ou continu, des discontinuités locales. Dans certains cas, la localisation des discontinuités de mouvement peut être facilitée par leur *coïncidence spatiale avec des frontières photométriques*.

La première partie de l'article (section 2) présente les méthodes utilisant la détection des frontières photométriques ou une pré-segmentation au sens photométrique, afin de limiter les zones d'application de la contrainte de lissage [Nagel'87 et 90], [Thiaville *et al.*'91], [Benois *et al.*'93]. Dans le cas général cependant, il est nécessaire d'utiliser l'information de mouvement seule. Ce cas est traité dans les sections 3 et 4 de l'article, consacrées respectivement aux approches « globale » et « locale ».

La deuxième partie (section 3) rapporte ainsi les méthodes visant à segmenter le champ estimé courant en régions homogènes au sens du mouvement. Cette segmentation fait généralement appel à un modèle paramétrique de mouvement [Murray et Buxton'87], [Bouhémy et François'93] et la contrainte de lissage est appliquée au sein de chaque région séparément.

La troisième partie (section 4) s'intéresse aux méthodes visant à estimer le champ de vitesses tout en détectant les ruptures locales de continuité de ce champ, afin d'inhiber le lissage dans les zones

susceptibles de contenir une discontinuité de mouvement. Cette inhibition peut être obtenue par introduction d'un « processus de ligne » binaire [Geman et Geman'84], [Konrad et Dubois'92], [Zhang et Hanauer'93 et 95], [Mémmin et Perez'95], grâce à l'utilisation d'un « estimateur robuste » [Geman et Reynolds'92], [Black et Rangarajan'93], [Gaidon *et al.*'93] ou dans un schéma de diffusion anisotrope [Nesi'93], [Proesmans *et al.*'94], [Deriche *et al.*'95]. Les ruptures locales de continuité du champ estimé courant peuvent également être détectées et prises en compte dans un schéma d'estimation faisant appel au filtrage de Kalman [Germain et Baudois'93b]. Nous citons aussi une méthode de détection de discontinuités de mouvement, préalable à toute estimation, à partir de la courbure de l'hypersurface de l'intensité dans l'espace-temps [Zetsche et Barth'91].

La dernière partie de l'article (section 5) est consacrée à la gestion des occultations. Le terme « zone d'occultation » comprend les parties de l'arrière-plan découvertes ou recouvertes au cours du mouvement des objets du premier plan. Dans le plan de l'image, les zones d'occultation sont adjacentes à des discontinuités de mouvement. Ce problème est donc lié à la présence de discontinuités de mouvement. Il a cependant été beaucoup moins abordé dans la littérature, probablement à cause de sa difficulté. En effet, il s'agit non seulement d'un non-respect de l'hypothèse de continuité, mais surtout d'une violation de l'hypothèse de conservation, puisque les motifs contenus dans les zones d'occultation apparaissent ou disparaissent d'une image à l'autre.

Précisons que les classifications proposées, en approches « globale » / « locale » et « photométrique » / « mouvement » ne sont pas dichotomiques, dans la mesure où certains auteurs utilisent conjointement les approches « photométrique » et « mouvement » [Black'92a], [Etoh et Shirai'93] ou « globale » et « locale » [Murray et Buxton'87], [Black et Jepson'94].

2. exploitation de la coïncidence entre discontinuités de mouvement et frontières photométriques

Les méthodes présentées dans cette section s'appuient sur une hypothèse plus ou moins implicite selon laquelle les objets en mouvement correspondent à des régions homogènes du point de vue photométrique ou, de manière duale, les discontinuités de mouvement coïncident avec certaines frontières photométriques. Tout comme dans le cas général, elles peuvent être subdivisées en deux catégories : une approche globale et une approche locale.

2.1. approche globale

L'approche globale consiste à pré-segmenter les images selon des critères photométriques (pré-segmentation statique) et à estimer le mouvement en appliquant la contrainte de continuité (lissage) séparément à l'intérieur de chaque Région Homogène au sens Photométrique (RHP). Un champ de vitesses initial peut être obtenu par application de la contrainte de conservation seule. Notamment, ce champ peut représenter la composante normale de la vitesse apparente, obtenue à partir de l'équation de contrainte de mouvement (1). L'estimation est alors régularisée, selon la formule (2) modifiée, afin d'appliquer la contrainte de lissage à l'intérieur de chaque RHP séparément [Thiaville *et al.*'91]. La modification consiste à inhiber le terme de lissage U_s lorsque les pixels voisins n'appartiennent pas à une même région. Le champ initial peut également être estimé avec une première régularisation qui ne tient pas compte des discontinuités [Black et Jepson'94]. Les auteurs proposent alors d'améliorer le résultat par identification d'un modèle paramétrique de mouvement pour chaque RHP séparément (cf section 3).

Une autre solution, pouvant s'inscrire dans le schéma pré-segmentation / estimation du mouvement par région, est présentée dans [Kottke et Sun'94]. La pré-segmentation consiste à agglomérer les points voisins ayant des intensités similaires. Chaque agglomérat de points (ang. *cluster*) ainsi obtenu est caractérisé par un vecteur de paramètres (position, intensité moyenne, moments d'inertie). L'estimation du mouvement d'un agglomérat est réalisée par une mise en correspondance itérative, consistant à minimiser une mesure de distance euclidienne entre les vecteurs de paramètres de cet agglomérat et de ses homologues potentiels dans l'image suivante. C'est donc une version améliorée de la mise en correspondance par blocs (ang. *block matching*), avec des « blocs » qui s'adaptent aux formes des objets. L'application de la contrainte de lissage au sein de chaque agglomérat est en revanche très sommaire : un même vecteur déplacement est affecté à tous les pixels formant l'agglomérat.

Dans le cas où l'objectif visé est la segmentation en régions homogènes au sens du mouvement, le champ de vitesses estimé peut être utilisé pour fusionner les RHP appartenant à un même objet [Benois *et al.*'93]. En effet, le nombre de régions obtenues par segmentation statique est généralement très supérieur au nombre réel d'objets en mouvement (fig. 2). Le champ de vitesses peut à son tour être ré-estimé au sein des Macro-Régions Homogènes au sens du Mouvement (MRHM). Le processus d'estimation / segmentation peut être itératif (fig. 3).

2.2. approche locale

2.2.1. détection explicite des frontières photométriques

La détection locale de contours photométriques est exploitée de deux manières différentes : soit pour ajuster une segmentation obtenue préalablement selon les critères de mouvement (cf section

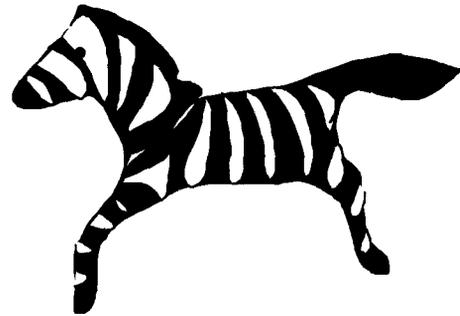


Figure 2. – Dans le cas de ce zèbre, le nombre de régions obtenues par segmentation statique, selon les critères photométriques, serait très supérieur au nombre de régions homogènes du point de vue du mouvement.

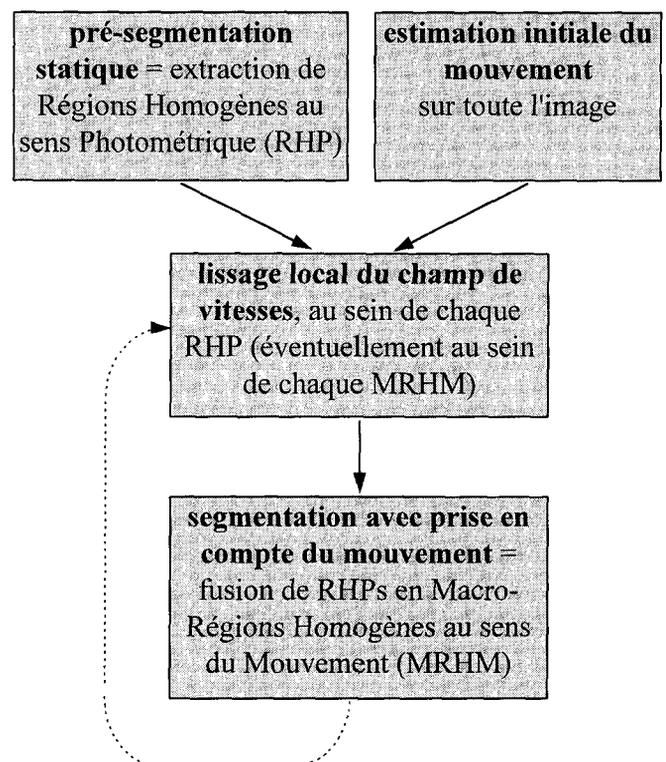


Figure 3. – Schéma de l'estimation / segmentation du mouvement avec pré-segmentation selon des critères photométriques.

3) [Diehl'91], [Bonnaud et Labit'95], [Zhang *et al.*'95], soit en interaction avec des processus de ligne qui modélisent la présence de discontinuités de mouvement et inhibent localement la contrainte de lissage [Konrad et Dubois'92], [Heitz et Bouthémy'89 et 93], [Zhang et Hanauer'93 et 95]. Ce dernier aspect sera développé plus loin (section 4), mais nous pouvons citer ici le cas, où les processus de ligne sont initialisés à l'aide des contours photométriques [Tian et Shah'94], [Devlamincq *et al.*'95].

On peut également citer une méthode particulière, dont l'objectif est l'estimation du mouvement de tous les contours, donc aussi des discontinuités de mouvement. Cette méthode, décrite dans

[Bouthémy'87, 88 et 89], exploite la dimension temporelle d'une séquence d'images, en modélisant un contour en mouvement par une portion de surface dans l'espace-temps. On considère un volume élémentaire spatio-temporel π . Deux hypothèses sont émises. Selon la première, ce volume ne contient pas de contour et les intensités des points appartenant à π suivent une loi gaussienne $N(c_0, \sigma^2)$, de moyenne c_0 et de variance σ^2 . Selon la deuxième, la présence d'un contour fait que π est divisé, par une surface, en deux parties π_1 et π_2 , où les intensités suivent respectivement les lois $N(c_1, \sigma^2)$ et $N(c_2, \sigma^2)$. La variance est supposée identique pour les deux hypothèses. La détection des contours et l'estimation de leurs mouvements se font simultanément, par un test du rapport de vraisemblance de ces deux hypothèses. Cette approche est décrite plus en détail au début de la section 3.3, à l'occasion de la présentation d'une autre méthode du même auteur, destinée à estimer le mouvement de régions.

2.2.2. lissage anisotrope

Au lieu de localiser explicitement les frontières photométriques, on peut effectuer un lissage anisotrope consistant à réduire la contribution des points potentiellement situés au-delà d'une discontinuité et à privilégier la contribution des points susceptibles d'appartenir au même objet que le point courant. Dans [Aisbett'89], la contribution de chaque point d'un voisinage du point courant, est pondérée par la probabilité d'appartenance au même objet. Cette probabilité, agissant selon l'auteur comme un détecteur implicite de contours, est inversement proportionnelle à la distance entre les pixels considérés et à la différence d'intensité lumineuse entre ces pixels.

Le lissage anisotrope peut être guidé par le gradient spatial [Nagel'87] ou spatio-temporel [Nagel'90] de l'intensité lumineuse. Il s'agit, dans ce cas, de lisser surtout la composante tangentielle (perpendiculaire au gradient) du vecteur vitesse. La composante normale, parallèle au gradient, est lissée d'autant moins que le module du gradient est grand. Le terme de lissage prend alors la forme suivante [Nagel'87] :

$$U_s = \iint \text{trace} \left[{}^t(\nabla \mathbf{v}) \mathbf{W} (\nabla \mathbf{v}) \right] dx dy \quad (4)$$

avec

$$\nabla \mathbf{v} = \begin{bmatrix} u_x & v_x \\ u_y & v_y \end{bmatrix}$$

et t désignant l'opération de transposition. L'anisotropie du lissage est obtenue grâce à la matrice de pondération \mathbf{W} dont les éléments sont calculés à partir des composantes du gradient de l'intensité lumineuse :

$$\mathbf{W} = \frac{1}{I_x^2 + I_y^2 + 2\gamma} \begin{bmatrix} I_y^2 + \gamma & -I_x I_y \\ -I_x I_y & I_x^2 + \gamma \end{bmatrix} \quad (5)$$

où γ est une constante permettant d'obtenir une matrice identité dans les zones d'intensité homogène (gradient nul). Une

approche similaire a été récemment proposée dans [Ghosal et Vanek'96]. Notons qu'un lissage anisotrope guidé par le gradient photométrique est également connu en stéréovision. Il sert alors à interpoler un champ épars de disparités en évitant de lisser les discontinuités de profondeur [Fua'91].

L'hypothèse de coïncidence des discontinuités avec des frontières photométriques correspond à une large classe de situations, mais ne peut pas s'appliquer à tous les cas. Notamment, le mouvement des objets « camouflés » (fig. 4), perceptible par l'œil humain, échappe totalement à cette approche.

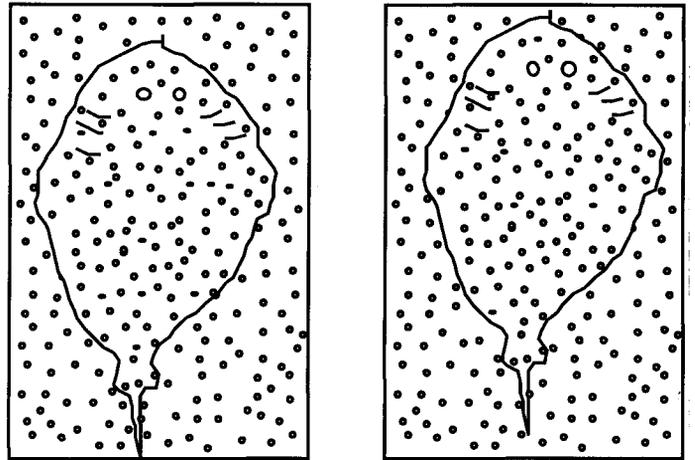


Figure 4. – Mouvement d'objet camouflé : sans contours tracés pour aider le lecteur, seul le mouvement permettrait de localiser cette « raie marbrée » camouflée sur « fond marin ».

3. segmentation du champ estimé courant en régions homogènes au sens du mouvement

La segmentation en régions homogènes au sens du mouvement met au premier plan la recherche des zones de continuité. Une fois ces régions délimitées, les discontinuités sont prises en compte naturellement, comme contours des régions.

La vitesse apparente étant une variable cachée (non observable), la segmentation du champ estimé courant s'appuie sur une information partielle de mouvement (généralement composante normale du vecteur vitesse), ainsi que sur un modèle paramétrique du mouvement. L'utilisation d'un tel modèle suppose la connaissance a priori qualitative du caractère du mouvement. En pratique, le choix du modèle est dicté par un compromis entre la capacité à

décrire une large gamme de mouvements et le besoin de limiter la complexité du modèle. Les modèles les plus fréquemment utilisés sont décrits dans la section 3.1.

Les tâches d'estimation des paramètres du modèle et de découpage en zones d'application de jeux de paramètres distincts, peuvent être accomplies l'une après l'autre ou simultanément. Ces deux catégories de méthodes, non disjointes, sont présentées dans les sections 3.2 et 3.3 respectivement. La section 3.4 est consacrée à deux méthodes particulières difficiles à classer dans les deux catégories précédentes.

3.1. modèles paramétriques de mouvement

Les différents modèles paramétriques de mouvement se caractérisent d'une part par leur complexité, d'autre part par leur signification physique. Dans [Sanson'93, 95a et b], on peut trouver une étude des problèmes soulevés par l'identification des modèles polynomiaux de degré n . Cependant, la complexité calculatoire et la difficulté d'interprétation physique limitent sensiblement le degré des modèles utilisés en pratique.

Le modèle paramétrique le plus fréquemment utilisé est le modèle linéaire 2D (affine) [Bouthémy'87], [François'91], [François et Bouthémy'91], [Bouthémy et François'93], [Benois *et al.*'93], [Black et Jepson'94], [Cloutier *et al.*'94], [Xiong et Graffigne'94], [Odobez et Bouthémy'94 et 95], [Bonnaud et Labit'95], [Duc *et al.*'95], [LeQuang *et al.*'95], [Zhang *et al.*'95] :

$$v = \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} p \\ q \end{bmatrix} + \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} = T + M(x - x_0) \quad (6)$$

où x_0 est un point de référence du mouvement (ex. centre de gravité de la région considérée), le vecteur T décrit une translation parallèle au plan de l'image et la matrice M peut décrire une combinaison de rotation, homothétie et déformation. Ce modèle à six paramètres permet de décrire la plupart des mouvements d'objets rigides (lorsque la géométrie du système d'acquisition permet de considérer que l'image est liée à la scène par une projection parallèle) et une partie des mouvements d'objets déformables. On peut également trouver dans la littérature un modèle linéaire simplifié, à quatre paramètres, avec $a = d$ et $c = -b$ [García-Garduno et Labit'93], [Xiong et Graffigne'94], [Bonnaud et Labit'95]. Ce modèle est en principe limité à la description des translations parallèles au plan de l'image et des rotations autour d'un axe perpendiculaire à ce plan. Le plus simple est le modèle constant qui suppose que tous les points d'une région effectuent un même mouvement de translation, décrit par le vecteur T seul, la matrice M étant alors nulle [Bouthémy'87], [Etoh et Shirai'93], [Black et Jepson'94], [Duc *et al.*'95], [LeQuang'95], [Chen *et al.*'95].

Outre ces modèles linéaires 2D, on trouve assez fréquemment un modèle de mouvement rigide en trois dimensions qui représente

une facette plane en translation et rotation [Murray et Buxton'87], [Diehl'91], [Chang *et al.*'93 et 94], [Black et Jepson'94], [Mitiche et Bouthémy'96]. Soit un point X de coordonnées ${}^t(X, Y, Z)$ dans l'espace 3D, situé sur une facette plane rigide. Le mouvement de la facette est décrit par :

$$V(X) = T + \Omega \times X \quad (7)$$

avec T et Ω représentant respectivement les composantes translationnelle et rotationnelle de la vitesse 3D du point X . En supposant que le système optique du capteur, caractérisé par une longueur focale l , réalise une projection perspective dont le centre optique est placé au point ${}^t(0, 0, 0)$ l'image du point X se trouve en :

$$x = \frac{-lX}{Z} = {}^t(x, y, -l) \quad (8)$$

Murray et Buxton démontrent que le mouvement dans le plan de l'image est alors décrit par :

$$\begin{aligned} u &= \beta_1 x + \beta_2 y - \beta_3 l + x(\beta_7 x + \beta_8 y)/l \\ v &= \beta_4 x + \beta_5 y - \beta_6 l + y(\beta_7 x + \beta_8 y)/l \end{aligned} \quad (9)$$

où $\beta_i (i = 1, \dots, 8)$ sont les paramètres du modèle. Notons que plusieurs modèles paramétriques, de différents niveaux de complexité, interviennent parfois dans un même algorithme. Un modèle simple intervient alors généralement dans une phase de segmentation initiale de l'image. Une amélioration de cette première segmentation, ou une description plus fine du mouvement de chaque région, est ensuite obtenue à l'aide d'un modèle plus complexe [Bouthémy'87], [Xiong et Graffigne'94], [LeQuang *et al.*'95]. Une autre démarche est proposée dans [Black et Jepson'94] : la complexité du modèle utilisé dépend ici de la taille de la région considérée. Les régions de petite taille sont traitées avec le modèle constant, les régions de taille moyenne avec le modèle affine 2D complet et les régions de grande taille avec le modèle 3D de facettes planes.

3.2. approche « séquentielle » : estimation des paramètres puis segmentation

Le trait d'union des méthodes présentées dans cette section est l'existence de deux opérations exécutées séquentiellement : identification du modèle puis segmentation. On peut cependant distinguer une sous-catégorie de méthodes qui se caractérise par un aspect hiérarchique, faisant appel à la notion de mouvement dominant (§ 3.2.1). Les autres méthodes (§ 3.2.2) s'efforcent de segmenter le flot optique en absence de mouvement dominant, en utilisant des techniques telles que le schéma classique de division / fusion (ang. *split and merge*) ou diverses techniques de classification.

3.2.1. approche hiérarchique

Les méthodes présentées dans cette section s'appuient, explicitement ou implicitement, sur l'hypothèse de présence d'un mouvement dominant. Ce terme désigne un mouvement pour lequel le support d'estimation des paramètres est nettement plus large que pour les autres mouvements observés dans l'image. C'est notamment le cas du mouvement de l'arrière-plan, dans les images prises par une caméra mobile ou fixe, lorsque les objets qui se déplacent au premier plan sont relativement petits. Dans ce cas, l'objectif de la segmentation est souvent de détecter ces objets. Le schéma adopté se décompose généralement comme suit. D'abord, le mouvement dominant est déterminé par identification d'un modèle global A pour toute l'image courante $I(t)$. Ensuite l'image précédente $I(t - dt)$ est compensée, c'est-à-dire chaque point x de cette image est déplacé à l'aide de la vitesse $v_A(x) = {}^t(u_A, v_A)$ déduite du jeu de paramètres A préalablement identifié. Puis l'erreur de compensation $DFD(x, A)$ (ang. *Displaced Frame Difference*) est calculée pour chaque pixel x :

$$DFD(x, A) = I(x, t) - I(x + v_A dt, t - dt) \quad (10)$$

Les objets dont le mouvement n'est pas conforme au modèle dominant A , sont en principe mal compensés. Leur détection est donc souvent faite par seuillage de l'erreur de compensation [Diehl'91], [Zhang *et al.*'95], mais on peut trouver des méthodes plus complexes, faisant par exemple appel à une modélisation markovienne [Odobez et Bouthémy'94].

L'aspect hiérarchique se manifeste dans le fait que le même schéma peut être appliqué à l'intérieur de chacun des objets ainsi détectés et ceci de manière récursive [Diehl'91], [Cloutier *et al.*'94]. Pour l'objet numéro k , on cherche le mouvement dominant A_k . Après la compensation à l'aide du mouvement estimé, on détecte les régions non conformes au modèle (correspondant généralement à des objets situés devant l'objet k) et ainsi de suite.

Le modèle correspondant au mouvement dominant d'une région R_k peut être identifié au sens des moindres carrés, comme c'est le cas dans [Diehl'91]. On minimise alors l'erreur globale de compensation, exprimée par l'énergie suivante :

$$U(A_k) = \sum_{x \in R_k} DFD^2(x, A_k) \quad (11)$$

La présence de plusieurs mouvements peut cependant sensiblement biaiser le résultat. Il est donc préférable d'utiliser un estimateur dit robuste [Black et Jepson'94], [Cloutier *et al.*'94], [Odobez et Bouthémy'94 et 95]. Les estimateurs robustes seront présentés dans la partie 4. Ici nous précisons seulement qu'il s'agit de remplacer la fonction quadratique, dans la formule (11), par une fonction qui limite l'influence des points aberrants (ang. *outliers*). Par points aberrants nous entendons les points situés hors du domaine de validité du modèle courant et donnant donc lieu à des erreurs de compensation importantes.

3.2.2. segmentation en l'absence de mouvement dominant

Dans cette catégorie, généralement, l'identification du modèle est d'abord faite localement, en chaque pixel. Les paramètres A_x du modèle sont estimés en utilisant un voisinage B_x du pixel x considéré, sans savoir si tous les points de ce voisinage appartiennent à un même objet. La segmentation consiste à regrouper les points connexes pour lesquels on a obtenu des jeux de paramètres similaires. Ceci peut être réalisé dans le schéma classique de division / fusion [Sanson'93]. Dans l'article cité, la phase de division utilise une structure d'arbre quaternaire (ang. *quadtree*) avec recouvrement. Il s'agit donc d'un découpage récursif en blocs carrés : chaque bloc est découpé en quatre s'il n'est pas homogène et ainsi de suite. La phase de fusion fait appel à la théorie des graphes. On définit un graphe orienté et pondéré d'adjacence de régions, où le poids d'un arc $R1 \rightarrow R2$ est égal à l'accroissement de l'erreur quadratique de compensation de la région $R1$ lorsque l'on remplace son modèle par celui de la région $R2$. Puis on détermine l'arbre couvrant minimum de ce graphe après avoir retenu, pour chaque couple de sommets, l'arc de coût minimum. Les régions sont fusionnées dans l'ordre imposé par cet arbre. L'algorithme est itératif : on procède à une nouvelle estimation des paramètres sur des supports déterminés par la segmentation précédente et ce jusqu'à convergence.

Le schéma division / fusion n'est pas la seule solution pour classer les points correspondant à des jeux de paramètres similaires. Nous avons trouvé dans la littérature une méthode consistant à utiliser la transformée de Hough. Ainsi, l'identification de mouvements multiples revient à chercher les pics dans l'espace de Hough [Zhang *et al.*'95]. L'identification de mouvements multiples peut également être réalisée à l'aide d'une classification selon les règles de la logique floue [Duc *et al.*'95], manipulant la notion de degré d'appartenance d'un même élément à différentes classes.

3.3. estimation et segmentation simultanées

La grande majorité des algorithmes permettant d'effectuer simultanément l'identification du modèle et le découpage en zones d'application de modèles à paramètres distincts, s'inscrit dans une approche probabiliste, bayésienne. Nous commençons cependant par citer une méthode qui se situe à la charnière avec les techniques présentées dans la section précédente.

3.3.1. test de vraisemblance

La méthode proposée dans [Bouthémy'87] utilise le schéma de division / fusion, comme les méthodes citées dans la section précédente, sans pour autant utiliser une phase d'identification préalable du modèle local pour chaque point. L'image est d'abord divisée en blocs carrés de taille 16×16 pixels, par exemple. Pour

chaque bloc R , on calcule le rapport de vraisemblance de deux hypothèses : H_0 selon laquelle le bloc est homogène au sens du mouvement, avec un modèle constant défini par une translation T_0 et H_1 selon laquelle le bloc est composé de deux parties $R1$ et $R2$ animées chacune d'un mouvement de translation différent T_1 et T_2 respectivement. A chaque hypothèse est associée une fonction de vraisemblance :

$$f(\varepsilon_A) = \prod_{\mathbf{x} \in R} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2} \varepsilon_A^2(\mathbf{x})\right) \quad (12)$$

où la variance σ^2 est supposée constante et identique pour les deux hypothèses. La variable ε_A , supposée indépendante d'un point à l'autre, est définie par :

$$\varepsilon_A(\mathbf{x}) = \nabla I(\mathbf{x}) \cdot \mathbf{v}_A + I_t. \quad (13)$$

Cette formule n'est rien d'autre que l'équation (1) avec la vitesse apparente \mathbf{v} remplacée par la vitesse déduite du jeu de paramètres A . Dans le cas présent, \mathbf{v}_A est égale à T_0 , T_1 ou T_2 respectivement. La variable ε_A mesure implicitement l'écart local entre le modèle et les données observées. La méthode citée consiste à rechercher d'abord les paramètres \hat{T}_0 , \hat{T}_1 et \hat{T}_2 minimisant le rapport logarithmique des fonctions de vraisemblance associées aux hypothèses H_0 et H_1 :

$$\chi(T_0, T_1, T_2) = \frac{1}{2\sigma^2} \sum_{\mathbf{x} \in R} \varepsilon_{T_0}^2(\mathbf{x}) - \frac{1}{2\sigma^2} \left[\sum_{\mathbf{x} \in R1} \varepsilon_{T_1}^2(\mathbf{x}) + \sum_{\mathbf{x} \in R2} \varepsilon_{T_2}^2(\mathbf{x}) \right] \quad (14)$$

et à comparer ensuite ce rapport à un seuil λ_{reg} prédéterminé. L'hypothèse H_0 l'emporte lorsque $\chi(\hat{T}_0, \hat{T}_1, \hat{T}_2) \leq \lambda_{reg}$. Dans le cas contraire, le bloc est divisé et la même procédure est appliquée à chacune de ses parties. La phase de fusion repose sur le même principe.

Une autre méthode d'estimation / segmentation de mouvements multiples à l'aide d'un test de vraisemblance est décrite dans [Jepson et Black'93]. Au lieu du schéma division / fusion cependant, cette méthode réalise une classification, avec un nombre de classes (mouvements différents) fixé à l'avance. La méthode a été conçue pour gérer le cas difficile des mouvements visibles par transparence.

3.3.2. approche bayésienne

Des distributions de probabilité semblables aux fonctions de vraisemblance (12) sont utilisées dans un cadre méthodologique différent : *bayésien*. Ce cadre, très général, s'applique aussi bien à la segmentation, à l'estimation du mouvement, qu'à bien d'autres problèmes en traitement d'images (et non seulement en traitement d'images). Le problème est posé de la manière suivante. On dispose d'un ensemble (champ) O de variables aléatoires appelées observations, à partir duquel on souhaite estimer un champ E de variables aléatoires appelées primitives ou étiquettes. Les

étiquettes peuvent être de nature très diverse : ici il s'agit avant tout de numéros de régions ou de leurs vecteurs de paramètres, mais il peut s'agir de vecteurs vitesse, comme nous le verrons plus loin, ou de bien d'autres entités. De même, les observations utilisées peuvent être variées. Dans le cas présent, les observations représentent le plus souvent les dérivées spatio-temporelles de la fonction intensité lumineuse. On cherche les primitives maximisant la probabilité globale a posteriori $p(E = e/O = o)$, où e et o sont respectivement des réalisations particulières des variables aléatoires E et O . On parle d'un estimateur au sens du *maximum a posteriori* (MAP). Conformément au théorème de Bayes :

$$p(E = e/O = o) = \frac{p(O = o/E = e)p(E = e)}{p(O = o)} \quad (15)$$

Le dénominateur étant une constante, maximiser $p(E = e/O = o)$ revient à maximiser la probabilité jointe $p(O = o/E = e)p(E = e) = p(O = o, E = e)$. Le premier terme, la probabilité a posteriori des observations $p(O = o/E = e)$, relie les primitives aux observations. Il constitue ainsi un terme d'attache aux données et prend généralement une forme gaussienne très similaire à (12) :

$$p(O = o/E = e) = \prod_{\text{image}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2} \varepsilon^2(\mathbf{x}, e)\right). \quad (16)$$

La fonction $\varepsilon(\mathbf{x}, e)$ mesure l'écart entre les valeurs estimées et les données observées. Cet écart dépend du champ des étiquettes courant : segmentation et modèle de mouvement identifié pour chaque région. La fonction $\varepsilon(\mathbf{x}, e)$ est définie soit par la formule (13) [François'91], [François et Bouthémy'91], [Bouthémy et François'93], [Etoh et Shirai'93], [Cloutier *et al.*'94], [Xiong et Graffigne'94], soit par l'erreur de compensation de mouvement *DFD* (10) [LeQuang *et al.*'95], soit encore par d'autres mesures proches des celles qui viennent d'être citées [Murray et Buxton'87], [Odobez et Bouthémy'94]. Au lieu de la distribution gaussienne, on peut trouver la distribution gaussienne généralisée [Stiller et Hürtgen'93], [Stiller'93 et 94].

Le terme de probabilité a priori $p(E = e)$ a un rôle régularisant. Il permet de mettre en équation les propriétés attendues du champ à estimer. Le formalisme *markovien* est particulièrement bien adapté à ce genre d'exercice. Soit S le support du champ à estimer, dans notre cas une grille bidimensionnelle de pixels. Chaque élément $s \in S$ est appelé site. On définit sur S un système de voisinage B . Le voisinage B_s du site s est généralement composé des 4 ou 8 sites les plus proches de s . Le champ E est de Markov pour le système de voisinage B s'il remplit deux conditions : 1) aucune réalisation n'est de probabilité nulle, 2) pour chaque site s_i , la loi de probabilité conditionnelle de son étiquette $p(e_i/e_j, s_j \in S - \{s_i\})$, sachant les étiquettes de tous les autres sites s_j de S , est la même que la loi $p(e_i/e_j, s_j \in B_{s_i})$ conditionnée uniquement par les étiquettes des sites voisins de s_i . D'après le théorème de Hammersley-Clifford, les probabilités associées à un champ de Markov suivent la distribution de Gibbs :

$$p(E = e) = \frac{1}{Z} \exp[-U(e)] \quad (17)$$

où Z est une constante de normalisation appelée fonction de partition. La fonction d'énergie $U(e)$ peut être décomposée en une somme de fonctions locales V_c appelées potentiels :

$$U(e) = \sum_{c \in C} V_c(e) \quad (18)$$

Chaque potentiel est associé à une clique $c \in C$, une clique étant un sous-ensemble de S composé soit d'un seul site, soit de sites qui sont tous mutuellement voisins, et C étant l'ensemble des cliques de S pour le système de voisinage B . Le potentiel associé à une clique c dépend uniquement des étiquettes portées par les sites de c . Maximiser la probabilité $p(E = e)$ revient à minimiser l'énergie $U(e)$. Pour formaliser les propriétés attendues du champ à estimer, il faut donc associer des potentiels bas aux configurations que l'on souhaite privilégier et des potentiels élevés aux configurations que l'on souhaite décourager. Dans le domaine de la segmentation en régions homogènes, on privilégie les cliques composées de sites ayant des étiquettes identiques. Soit une clique c binaire, c'est-à-dire composée de deux sites s_i et s_j . Le potentiel V_c associé à cette clique, noté aussi V_{ij} , est souvent défini par la formule suivante [François'91], [François et Bouthémy'91], [Bouthémy et François'93], [Etoh et Shirai'93], [Cloutier *et al.*'94], [Xiong et Graffigne'94] :

$$V_{ij} = \mu(1 - \delta_{e_i, e_j}) \quad (19)$$

où e_i et e_j sont les étiquettes respectives des sites s_i et s_j , δ symbolise la fonction delta de Kronecker prenant la valeur 1 lorsque $e_i = e_j$ et la valeur 0 dans le cas contraire, tandis que $\mu > 0$ est un paramètre prédéterminé pénalisant l'introduction d'une discontinuité (configuration $e_i \neq e_j$). Un choix similaire consiste à remplacer le potentiel nul par un potentiel négatif [Murray et Buxton'87], [Chang *et al.*'93 et 94] :

$$V_{ij} = \mu(1 - 2\delta_{e_i, e_j}) = \begin{cases} -\mu & \text{si } e_i = e_j \\ \mu & \text{sinon} \end{cases} \quad (20)$$

L'estimation du champ de segmentation est parfois couplée avec l'estimation d'un autre champ. Le plus souvent, il s'agit du champ de vitesses, pour lequel les primitives sont les vecteurs vitesse. Les potentiels favorisant un champ de vitesses lisse par morceaux sont généralement de la forme suivante [Stiller et Hürtgen'93], [Stiller'93 et 94], [Chang *et al.*'94] :

$$V_{ij} = \|\tilde{\mathbf{v}}_i - \tilde{\mathbf{v}}_j\|^2 \delta_{e_i, e_j} \quad (21)$$

où $\tilde{\mathbf{v}}_i$ et $\tilde{\mathbf{v}}_j$ sont les valeurs estimées courantes des vitesses apparentes associées aux sites s_i et s_j respectivement. Le potentiel croît rapidement avec la différence de vitesses entre les sites voisins ce qui privilégie le lissage. Mais le lissage est suspendu lorsque les deux sites n'appartiennent pas à une même région de la segmentation courante.

Les trois formules citées ci-dessus ne constituent bien évidemment pas une liste exhaustive de potentiels utilisés par différents

auteurs, en fonction des propriétés qu'il souhaitent privilégier. Nous citerons encore seulement une propriété, sans doute trop peu exploitée : la *continuité temporelle*. On favorise alors les configurations dans lesquelles chaque site conserve au cours du temps la même étiquette [Murray et Buxton'87], [Chang *et al.*'93], [Stiller'94].

Revenons maintenant à notre problème bayésien de départ (15). La distribution a posteriori des observations et la distribution a priori des étiquettes étant exponentielles, le logarithme de la distribution jointe $p(O = o/E = e)p(E = e) = p(O = o, E = e)$ est, au signe près, de la forme (2) avec le terme d'attache aux données et le terme régularisant définis par les formules suivantes :

$$U_d(o/e) = \sum_{\text{image}} \varepsilon^2(\mathbf{x}, e) \stackrel{\text{df}}{=} \sum_{s \in S} V_s^2(e) \quad (22)$$

et

$$U_s(e) = \sum_{c \in C} V_c(e)$$

On néglige ici les termes considérés comme constants, notamment les logarithmes de la variance σ^2 et de la fonction de partition Z . Connaissant ce résultat, il n'est pas surprenant que certains auteurs posent le problème d'estimation bayésienne au sens du MAP, directement en termes de minimisation d'une énergie gibbienne, avec un (ou plusieurs) terme(s) d'attache aux données et un (ou plusieurs) terme(s) régularisant(s) [François'91], [François et Bouthémy'91], [Bouthémy et François'93], [Chang *et al.*'94], [Odobez et Bouthémy'94]. Les fonctions d'erreur $\varepsilon(\mathbf{x}, e)$ sont alors considérées comme des potentiels $V_s(e)$ associés aux cliques unaires (composées chacune d'un seul site s). L'énergie globale est généralement minimisée à l'aide d'un algorithme déterministe du type ICM (ang. *Iterated Conditional Modes*) [François'91], [François et Bouthémy'91], [Bouthémy et François'93], [Chang *et al.*'93 et 94], [Stiller et Hürtgen'93], [Stiller'93 et 94], [Cloutier *et al.*'94], [Xiong et Graffigne'94] ou HCF (ang. *Highest Confidence First*) [Odobez et Bouthémy'94]. Ces algorithmes permettent d'éviter la lenteur de convergence des algorithmes stochastiques tels que le recuit simulé [Murray et Buxton'87], mais conduisent parfois à une solution sous-optimale (minimum local). Pour contourner ce piège, certains adoptent un schéma multi-résolution [Chang *et al.*'94], [Odobez et Bouthémy'94].

3.4. méthodes particulières

3.4.1. utilisation d'un contour déformable

La méthode consiste à rechercher le contour extérieur (fermé) d'un objet en mouvement à partir d'une frontière initiale (très approximative), placée à l'intérieur de l'objet [Schnörr'92]. L'auteur indique que le placement de la frontière initiale demande une connaissance a priori de la structure qualitative du champ, permettant de distinguer approximativement l'objet mobile du fond. Par exemple, on considère un champ globalement divergeant (cas d'une caméra qui avance ou effectue un zoom), dans lequel l'objet

en mouvement effectuée une translation parallèle au plan de l'image. L'auteur ne donne pas plus de détails quant à cette détection initiale, mais on peut très bien imaginer effectuer l'identification du mouvement dominant, suivie d'une détection des régions mal compensées, par une des techniques exposées dans la section 3.2.1.

L'algorithme est itératif et alterne l'estimation du flot optique et la déformation du contour. Soit R_0^i la région délimitée par la frontière détectée, à l' i -ème itération. Le flot optique est estimé par minimisation du critère de Horn et Schunck (2), (3) à l'intérieur et à l'extérieur de R_0^i séparément. La frontière de R_0^i est ensuite déformée, afin de minimiser une fonction de coût qui implique les gradients de l'intensité et les gradients des vitesses, mesurés à l'intérieur et à l'extérieur de R_0^i . On obtient ainsi R_0^{i+1} , on recalcule les vitesses à l'intérieur et à l'extérieur de R_0^{i+1} et ainsi de suite.

3.4.2. extraction d'harmoniques

La méthode d'estimation du mouvement à laquelle est consacrée cette section [Chen *et al.*'95] représente une approche connue depuis une décennie sous le nom de l'approche fréquentielle ou spatio-temporelle. Elle consiste à travailler dans l'espace des fréquences spatiales et temporelles, ce qui nécessite l'utilisation de toute une séquence d'images au lieu d'estimer le mouvement entre deux images consécutives seulement (comme c'est souvent le cas dans les autres approches). La plupart des méthodes spatio-temporelles s'appuient sur une hypothèse selon laquelle le mouvement est constant dans la fenêtre spatio-temporelle considérée [Jacobson et Wechsler'87], [Heeger'87]. Cette hypothèse revient à négliger la présence d'éventuelles discontinuités et ne permet pas d'estimer des mouvements multiples.

La méthode présentée dans [Chen *et al.*'95] s'affranchit partiellement de cette limitation. L'estimation concerne le mouvement de la partie « toujours visible » de chaque région, c'est-à-dire celle qui ne subit pas d'occultation. On utilise les notations et hypothèses suivantes. Une séquence de T images $\{I(x, y; t), t = 1, 2, \dots, T\}$ est considérée comme une superposition de N régions en mouvement, N étant à estimer. Sur une courte période, le mouvement de chaque région R_k est assimilé à une translation $T_k = {}^t(p_k, q_k)$. On considère une séquence de T transformées de Fourier 2D spatiales $F(\omega_x, \omega_y; t)$ des images consécutives :

$$\{F(\omega_x, \omega_y; t) = \sum_{k=1}^N F_k(\omega_x, \omega_y) e^{-j(p_k \omega_x + q_k \omega_y)t} + B(\omega_x, \omega_y; t), t = 1, 2, \dots, T\} \quad (23)$$

où $B(\omega_x, \omega_y; t)$ est un terme de bruit. Pour les fréquences spatiales (ω_x, ω_y) fixées, la série temporelle $F(\omega_x, \omega_y; t)$ est une somme de N harmoniques, dont les fréquences et les amplitudes s'expriment respectivement par :

$$\omega_k = p_k \omega_x + q_k \omega_y \text{ et } |F_k(\omega_x, \omega_y)|$$

avec $k = 1, 2, \dots, N$. L'estimation des mouvements des N régions est alors faite par une recherche des plus grands maxima locaux du spectrogramme associé à $\{F(\omega_x, \omega_y; t)\}$:

$$G(\omega_x, \omega_y; \omega) = \frac{1}{T} \left| \sum_{t=0}^{T-1} F(\omega_x, \omega_y; t) e^{-j\omega t} \right|^2 \quad (24)$$

L'algorithme de recherche des maxima du spectrogramme n'est pas itératif, ce qui est intéressant du point de vue du temps de calcul. Un autre atout de cette méthode est, d'après les auteurs, sa capacité à traiter des cas où les discontinuités sont nombreuses et rapprochées. Ce propos est illustré par une estimation satisfaisante du mouvement d'une grille fine et transparente, sur un fond lui aussi en mouvement. Rappelons cependant que la méthode est limitée aux mouvements translationnels uniformes.

4. prise en compte des ruptures locales de continuité du champ estimé courant

Les méthodes présentées dans cette partie de l'article utilisent une détection explicite ou implicite des zones susceptibles de contenir une discontinuité de mouvement. Excepté une méthode [Zetzsche et Barth'91], toutes les autres recherchent les discontinuités dans le champ estimé courant, ce qui veut dire que le calcul du flot optique et la détection des discontinuités sont réalisées conjointement. La section 4.1 rend compte de la diversité des approches proposées par différents auteurs, chacune des méthodes citées ayant cependant fait, du moins jusqu'à présent, relativement peu d'adeptes. Les deux sections suivantes en revanche, présentent des approches connaissant aujourd'hui un grand succès. La section 4.2 est consacrée à l'utilisation de processus de ligne. La section 4.3 introduit la notion d'estimateur robuste. La section 4.4 enfin, évoque l'utilisation de la diffusion anisotrope dans le cadre de l'estimation du mouvement. La diffusion anisotrope, actuellement moins répandue que les processus de ligne ou les estimateurs robustes, semble cependant connaître un succès croissant, les trois approches ayant d'ailleurs des points communs.

4.1. méthodes diverses

Cette section présente brièvement des exemples choisis de méthodes n'ayant pas nécessairement de traits communs évidents. Cette absence de caractéristiques communes témoigne de la diversité des moyens utilisés, historiquement et aujourd'hui encore, pour caractériser la présence des ruptures locales de continuité du flot optique.

4.1.1. détection des discontinuités par convolution avec un masque

Une première idée consiste à traiter le flot optique comme un couple d'images, chacune représentant respectivement la composante horizontale ou verticale des vecteurs vitesse. On peut alors appliquer une technique de détection de contours classique, connue en traitement d'image. Ainsi les discontinuités du champ de vitesses peuvent être détectées en localisant les passages par zéro du laplacien appliqué aux composantes verticale et horizontale de ce champ [Thompson *et al.*'85]. Cette opération, dans l'article cité, est un premier pas vers l'analyse de la structure de la scène filmée (ang. *structure from motion*), visant à distinguer les points appartenant aux objets situés en avant-plan, des points appartenant aux objets situés en arrière-plan.

On peut également chercher les maxima locaux des gradients spatiaux du champ de vitesses. Le calcul des gradients se fait alors par convolution de chacune des images-composantes du flot optique avec le masque de Prewitt par exemple [Schunck'89]. Dans la méthode citée, on effectue, après cette détection, un lissage du champ de vitesses par calcul variationnel, faisant appel à une approximation du laplacien. La présence des discontinuités pose le problème de gestion des effets de bord. En effet, il est difficile d'implanter des approximations du laplacien sur des supports à géométrie variable, permettant d'exclure du support les points appartenant à une discontinuité de forme quelconque. Pour calculer une nouvelle valeur lissée de la vitesse en un point donné, l'auteur propose, dès lors que le support (carré) de calcul du laplacien contient une discontinuité, de recopier la vitesse du pixel central du support sur tous les points de discontinuité détectés au sein du support. L'algorithme est itératif : on refait la détection des discontinuités sur le champ lissé, puis on lisse à nouveau en prenant en compte les discontinuités ainsi corrigées, etc. Notons que les effets de bord peuvent être pris en compte dans un schéma plus élaboré, proposé dans [Schnörr'91], où l'adaptation aux formes des discontinuités est obtenue grâce à l'utilisation d'éléments finis triangulaires.

La méthode proposée dans [Murray et Williams'86] n'utilise pas une convolution au sens propre du terme. Il y a cependant une certaine ressemblance, dans la mesure où il s'agit également de « promener » une fenêtre sur toute l'image et, pour chaque position de la fenêtre, attribuer à son pixel central une valeur, fonction des points voisins. On utilise le même modèle paramétrique de mouvement que dans [Murray et Buxton'87] (cf section 3.1), mais le but est de détecter les frontières locales entre les zones d'application de jeux de paramètres distincts et non de segmenter en régions homogènes au sens du modèle. Le jeu de paramètres A_x pour le point x est calculé à l'aide des points de la fenêtre centrée sur ce point. On suppose que les points situés sur un même objet en mouvement donnent des vecteurs de paramètres A_x similaires et qu'une discontinuité de mouvement se caractérise par un changement brusque de la norme de ce vecteur $\|A_x\|$.

4.1.2. autres méthodes de détection explicite des discontinuités de mouvement

Nous présentons ici trois autres méthodes de détection explicite des discontinuités de mouvement. Les deux premières, comme la méthode d'extraction d'harmoniques (cf § 3.4.2), exploitent la dimension temporelle d'une séquence d'images. La première analyse le signal d'innovation d'un filtre de Kalman. La deuxième s'intéresse à la courbure de l'hypersurface 4D espace-temps-intensité. La troisième méthode est spécifique à l'estimation du mouvement par mise en correspondance de blocs et exploite essentiellement les propriétés de la surface de corrélation (ce terme sera expliqué plus loin).

Filtrage de Kalman.

L'approche présentée dans [Germain et Baudois'93a et b], [Germain et Skordas'94], [Germain'95] consiste à attacher un filtre de Kalman temporel à chaque pixel de l'image. Le filtrage de Kalman offre un mécanisme de pondération dynamique entre le lissage (intégration des observations « passées ») et la prise en compte des nouvelles observations. Cette pondération est assurée au travers du gain du filtre, recalculé à chaque pas, en fonction de la covariance du bruit d'évolution et de la covariance du bruit d'observation. L'idée consiste à jouer sur la covariance du bruit d'évolution, afin de réduire le lissage lorsque la probabilité d'une discontinuité augmente, voire à réinitialiser le filtre lorsque l'on détecte une discontinuité potentielle. La variable d'état du filtre est constituée de la vitesse instantanée du pixel considéré, ainsi que ses dérivées successives. L'ordre de ces dérivées dépend de l'application et n'est pas fixé a priori. En pratique, la vitesse est souvent supposée comme constante entachée d'un bruit blanc, gaussien, centré. D'éventuelles accélérations sont alors prises en compte dans le terme de bruit d'évolution qui permet au filtre de suivre l'évolution d'un vecteur vitesse variant lentement dans le temps. L'observation est la composante normale du vecteur vitesse recherché, calculée à l'aide des gradients spatio-temporels de l'intensité.

Tant que le pixel donné fait partie d'un même objet en mouvement (fig. 5a), le filtre effectue un lissage temporel. Lorsqu'une discontinuité « passe » par le pixel considéré (fig. 5b), le filtre est réinitialisé conditionnellement compte tenu des mouvements estimés auparavant dans un voisinage temporel ou spatio-temporel de ce pixel. La détection des discontinuités potentielles est faite à partir des statistiques du signal d'innovation fourni par le filtre, au moyen d'un test du χ^2 . Les différentes versions de cette approche se distinguent surtout par la stratégie de réinitialisation.

Courbure.

La méthode proposée dans [Zetsche et Barth'91] s'appuie sur les techniques de la géométrie différentielle. Le signal 3D $I(x, y, t)$ est considéré comme une hypersurface dans un espace 4D $S(x, y, t) = \{(x, y, t, \eta) : \eta = I(x, y, t)\}$. La courbure de cette hypersurface est intrinsèquement tridimensionnelle pour les frontières de mouvement (bidimensionnelle pour les contours

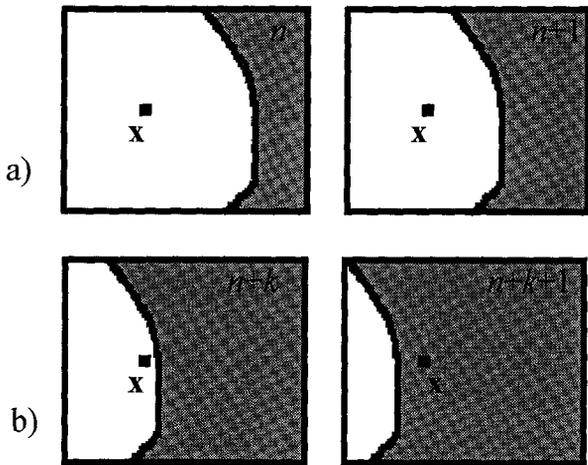


Figure 5. – Notion de continuité temporelle pour un pixel fixé : (a) entre les images n et $n + 1$, le pixel x reste au sein d'un même objet, il y a continuité, (b) entre les images $n + k$ et $n + k + 1$ le pixel x change d'objet (on dit qu'il a été traversé par une discontinuité), le mouvement estimé pour certains points de son voisinage correspond au deuxième objet et peut être utilisé pour réinitialiser le filtre de Kalman attaché au pixel x .

situés à l'intérieur des objets et unidimensionnelle si ces contours sont rectilignes). Les auteurs proposent donc de déterminer la courbure gaussienne de l'hypersurface considérée :

$$K = -\frac{\det H}{(I_x^2 + I_y^2 + I_t^2 + 1)^{5/2}} \quad (25)$$

Pour ce faire, ils calculent la matrice hessienne H de $I(x, y, t)$ et exploitent soit son déterminant :

$$\det H = I_{xx}I_{yy}I_{tt} + 2I_{xy}I_{xt}I_{yt} - I_{xx}I_{yt}^2 - I_{yy}I_{xt}^2 - I_{tt}I_{xy}^2 \quad (26)$$

soit la plus petite de ses valeurs propres. Le principal atout de cette méthode est de localiser les zones de discontinuité de mouvement avant toute tentative d'estimation de mouvement. Elle présente cependant l'inconvénient de ne pas permettre la localisation des frontières de mouvement rectilignes, qui possèdent les mêmes propriétés de courbure bidimensionnelle que les contours curvilignes situés à l'intérieur des objets en mouvement.

Propriétés de la surface de corrélation

L'estimation du mouvement d'un bloc par mise en correspondance consiste à calculer la corrélation, ou plus exactement l'intercorrélation, entre ce bloc, extrait d'une image, et des blocs de la même taille, placés au sein d'une fenêtre de recherche dans une autre image. Les valeurs de corrélation calculées pour l'ensemble des blocs de la fenêtre de recherche peuvent être représentées par une surface dont le pic le plus élevé indique le déplacement estimé. D'après [Black et Anandan'90b], lorsque le bloc contient une discontinuité de mouvement, la surface d'intercorrélation présente deux pics, dont les hauteurs relatives sont d'autant plus proches que la discontinuité passe près du milieu du bloc. Une première mesure de présence d'une discontinuité est donc le rapport de ces hauteurs relatives. Une deuxième mesure est proportionnelle

à la distance entre cette surface et la surface d'autocorrélation du bloc considéré. La surface d'autocorrélation est construite de la même manière que la surface d'intercorrélation, mais en calculant la corrélation entre le bloc considéré et l'image de laquelle il a été extrait. Ces deux surfaces devraient être très similaires en absence de discontinuité. Enfin, les auteurs proposent de construire l'histogramme des déplacements estimés pour les pixels d'un voisinage et de sélectionner les deux pics les plus élevés. Chaque pic est censé correspondre au mouvement d'un objet présent dans le voisinage concerné. La troisième mesure de présence de discontinuité est le rapport des hauteurs du deuxième et du premier pic. Ce rapport atteint un maximum proche de 1.0 lorsque le point central se situe approximativement sur la frontière entre les deux objets. Ces trois mesures sont utilisées conjointement avec une hypothèse selon laquelle une discontinuité de mouvement n'est pas un point isolé, mais forme une courbe dans l'espace. On cherche à approcher cette courbe par un modèle déformable de type « snake ». L'énergie externe qui sert à déformer le « snake » est calculée à l'aide des trois mesures évoquées ci-dessus.

4.1.3. prise en compte des discontinuités de mouvement sans détection explicite

Certains auteurs tentent d'appliquer la contrainte de lissage de manière sélective, seulement pour les points appartenant à une même région en mouvement, sans pour autant détecter explicitement les discontinuités de mouvement. Nous donnons ici deux exemples d'approches, très différentes l'une de l'autre. La première, purement heuristique, s'appuie sur des propriétés de la mise en correspondance par blocs. La deuxième fait appel à la notion de droite de contrainte de mouvement qui est une représentation graphique de l'équation (1).

Lissage anisotrope guidé par l'erreur de ressemblance.

La méthode résumée ci-dessous [Zheng et Blostein'92] consiste à estimer un champ initial par mise en correspondance de blocs et à effectuer ensuite un lissage anisotrope de ce champ (cf § 2.2.2). Dans cette deuxième phase, on cherche à minimiser une fonctionnelle d'énergie de type (2), avec :

$$\begin{aligned} U_d(v) &= \iint \|v(x) - d(x)\|^2 dx dy \\ U_s(v) &= \iint \|v(x) - \bar{v}(x)\|^2 dx dy \end{aligned} \quad (27)$$

$d(x)$ étant une vitesse initiale (plus exactement déplacement initial, ce qui est équivalent si l'on pose $dt = 1$), estimée par mise en correspondance de blocs et $\bar{v}(x)$ étant une vitesse moyenne calculée sur un voisinage $B(x)$:

$$\bar{v}(x) = \frac{\sum_{x_i \in B(x)} c(x_i) v(x_i)}{\sum_{x_i \in B(x)} c(x_i)} \quad (28)$$

où la contribution de chaque point x_i est pondérée par sa mesure de confiance $c(x_i)$. Cette mesure de confiance est définie de

manière à prendre des valeurs élevées pour les blocs contenant un motif discriminant. Elle prend des valeurs faibles aussi bien pour les blocs contenant une discontinuité de mouvement que pour les blocs situés dans une zone homogène du point de vue photométrique. Le calcul de la mesure de confiance utilise la valeur de corrélation entre le bloc centré sur x_i dans l'image précédente et son homologue trouvé dans l'image courante, ainsi que la variance de cette corrélation au sein de la fenêtre de recherche.

Regroupement d'intersections des droites de contrainte de mouvement

Une droite de contrainte de mouvement (fig. 6) est la représentation graphique de l'équation (1), dans les coordonnées du vecteur vitesse (u, v) . En d'autres termes, une droite de contrainte représente, dans l'espace des vitesses, l'ensemble des vitesses qui satisfont l'équation (1). Cette représentation met en évidence un degré d'indétermination. L'indétermination peut être levée en utilisant les droites de contrainte de plusieurs (au moins deux) points animés d'un même mouvement, lorsque l'orientation du gradient spatial varie entre ces points. Idéalement, les droites correspondant aux points animés d'un même mouvement de translation possèdent une seule intersection qui détermine de

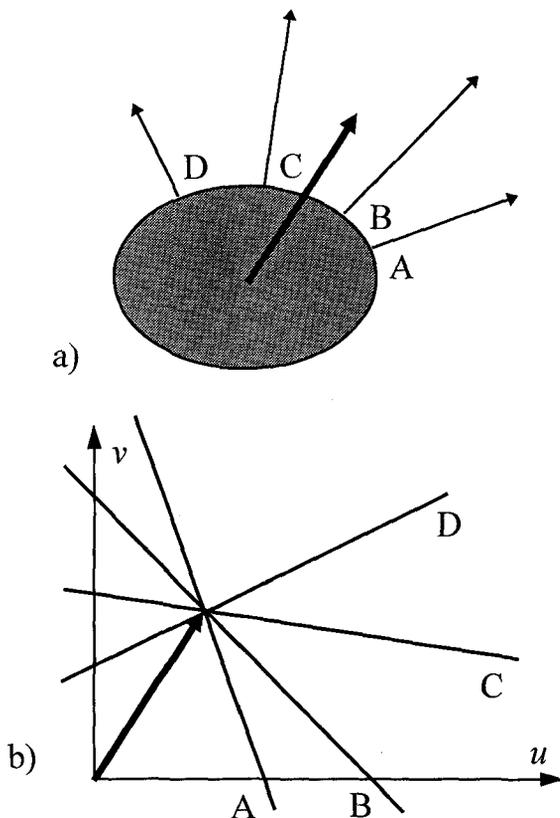


Figure 6. – Détermination du mouvement à l'aide de droites de contrainte de plusieurs points : (a) un objet en mouvement de translation (la flèche épaisse correspond à la « vraie » vitesse de l'objet; les flèches fines représentent la composante « normale » du vecteur vitesse pour les points A, B, C, et D), (b) les droites de contrainte correspondant aux points A, B, C, et D.

manière unique les deux composantes du vecteur vitesse (fig. 6). Dans les cas réels cependant, même si le mouvement est purement translationnel, les points d'intersection sont dispersés, à cause du bruit et des variations insuffisantes de l'orientation du gradient spatial.

Si le nuage des points d'intersection obtenu correspond à un seul mouvement, l'identification du « vrai » mouvement peut être faite par les moindres carrés ou par leur variante TLS (ang. *Total Least Squares*) [Chu et Delp'89]. La prise en compte des discontinuités s'appuie sur l'hypothèse selon laquelle un seul mouvement donne un nuage compact de points d'intersection, tandis que la présence de plusieurs mouvements distincts se manifeste par l'existence de plusieurs amas de points d'intersection [Black et Jepson'94]. Pour déterminer le mouvement d'un pixel donné, en évitant le biais provoqué par la contribution des pixels situés au-delà d'une discontinuité (dont on ne connaît pas la position exacte), [Schunck'89] propose d'estimer le mouvement localement dominant. Pour ce faire, il utilise une analyse unidimensionnelle des intersections de la droite de contrainte associée au pixel considéré, avec les droites des pixels voisins. Cette analyse consiste à chercher le plus petit intervalle contenant la moitié de points d'intersection. Notons une proposition intéressante d'éliminer les intersections des droites quasiment parallèles, en prenant deux à deux les équations de contrainte de différents pixels et en comparant à un seuil le déterminant de chaque système ainsi obtenu [Tistarelli'95].

4.2. processus de ligne

Depuis leur introduction par [Geman et Geman'84], dans le domaine de la restauration d'images, les « processus de ligne » se sont répandus également en estimation du mouvement [Murray et Buxton'87], [Konrad et Dubois'92], [Black'92a], [Heitz et Bouthémy'89 et 93], [Zhang et Hanauer'93 et 95], [Tian et Shah'94]. Cette approche est étroitement liée à l'estimation bayésienne et, plus précisément, à la modélisation markovienne de la structure a priori du champ à estimer. Le formalisme correspondant a été résumé dans la section 3.3.2. Ici nous allons nous concentrer sur les différentes façons de définir les fonctions de potentiel dans les méthodes faisant appel aux processus de ligne.

Soit un champ à estimer (champ principal) supposé continu par morceaux. Un processus de ligne (ang. *line process*) est une variable aléatoire cachée permettant, lorsqu'elle est activée, d'inhiber localement le lissage. Cette variable est censée marquer les discontinuités présentes entre des sites du champ principal. Elle est associée à un champ dual (champ de discontinuités) dont les sites d_m sont placés entre les sites pixels s_i du champ principal (fig. 7). Les deux champs sont estimés simultanément, au cours d'un même processus d'optimisation.

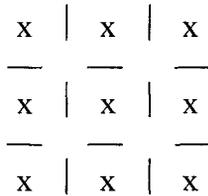


Figure 7. – Représentation graphique de la maille de sites pixels (x) et sites discontinuités (traits horizontaux et verticaux).

4.2.1. potentiels associés au champ principal

Soient X et L les variables aléatoires associées respectivement au champ principal et au champ dual. La variable X représente l'écart local entre le champ estimé courant et son modèle. Par exemple, pour un modèle du champ de vitesses (champ principal) localement constant et pour une clique formée de deux sites pixels s_i et s_j cet écart est généralement calculé comme la différence entre les vitesses estimées pour les sites s_i et s_j : $x_{ij} = \| \tilde{v}(s_i) - \tilde{v}(s_j) \|$. Pour un voisinage d'ordre 2, on peut avoir des cliques formées par un triplet de sites pixels s_i, s_j et s_k avec s_i placé entre s_k et s_j . La variable X est alors définie par : $x_{ijk} = \| 2 \tilde{v}(s_i) - \tilde{v}(s_j) - \tilde{v}(s_k) \|$ [Heitz et Bouthémy'89].

La variable L , processus de ligne, est généralement binaire : $l_{ij} = 1$ et $l_{ij} = 0$ correspondent respectivement à la présence et à l'absence de discontinuité entre les sites pixels s_i et s_j [Murray et Buxton'87], [Heitz et Bouthémy'89], [Black'92a], [Konrad et Dubois'92], [Zhang et Hanauer'93 et 95], [Tian et Shah'94]. On peut également rencontrer des processus de ligne trinaires [Heitz et Bouthémy'93] ou continus $0 \leq l_{ij} \leq 1$ [Geman et Reynolds'92]. Dans le cas trinaire, la variable L prend les valeurs $-1, 0$ ou $+1$. La présence d'une discontinuité est indiquée par la valeur absolue de l_{ij} ($|l_{ij}| = 1$). Le signe sert à gérer les occultations (cf partie 5) et indique quel côté de la discontinuité correspond à une surface occultée (second plan).

Hormis le cas des processus de ligne trinaires, où la formule est plus complexe, le potentiel associé à une clique formée de deux sites pixels s_i et s_j , est donné par la fonction suivante :

$$V_{ij} = (1 - l_{ij}) x_{ij}^2 + \psi(l_{ij}) \quad (29)$$

Pour les processus de ligne binaires, le second terme est, lui aussi, binaire : $\psi(l_{ij}) = \beta l_{ij}$, avec la constante β strictement positive. Le potentiel V_{ij} représente le coût associé à la clique considérée lorsque le champ estimé s'écarte localement du modèle lisse. En absence de discontinuité, le coût croît quadratiquement avec la différence entre les vitesses estimées pour les sites considérés, ce qui favorise le lissage du champ. Le terme quadratique peut cependant être annulé en activant le processus de ligne placé entre les sites pixels considérés ($l_{ij} = 1$). Il faut donc associer une pénalité à l'activation du processus de ligne, pour éviter d'introduire des discontinuités partout. C'est le rôle du terme βl_{ij} . Lorsque la variable L est continue, le terme constant βl_{ij}

est remplacé par une fonction continue, par exemple $\psi(l_{ij}) = (\sqrt{l_{ij}} - 1)^2$ [Black et Rangarajan'93].

4.2.2. potentiels associés au champ dual

Rappelons que le potentiel V_{ij} était associé à une clique constituée de sites pixels. On peut de même associer des potentiels à des cliques formées de sites discontinuités. Dans la littérature, on trouve essentiellement deux catégories de potentiels V_d associés au champ dual, modélisant respectivement : les propriétés géométriques des discontinuités et les interactions entre les discontinuités de mouvement et les frontières photométriques. Dans le premier cas, chaque clique est composée de plusieurs sites discontinuités. Dans le deuxième cas, on s'intéresse aux cliques à un seul site.

Le potentiel utilisé pour favoriser la coïncidence entre les discontinuités de mouvement et des frontières photométriques peut être binaire ou continu. Un potentiel binaire [Heitz et Bouthémy'93] est utilisé lorsque les frontières photométriques sont au préalable détectées explicitement et marquées par une variable binaire η ($\eta_m = 1$ si un contour passe par le site discontinuité d_m placé entre les sites pixels s_i et s_j) :

$$V_{dm} = \beta_d (1 - \eta_m) l_{ij} \quad (30)$$

Lorsque les frontières photométriques ne sont pas détectées explicitement, [Konrad et Dubois'92] et [Zhang et Hanauer'93 et 95] proposent d'utiliser un potentiel inversement proportionnel au carré du gradient d'intensité entre les sites s_i et s_j :

$$V_{dm} = \beta_d l_{ij} / (I_i - I_j)^2 \quad (31)$$

Le potentiel censé favoriser ou défavoriser certaines configurations géométriques des discontinuités met en interaction des sites discontinuités voisins. La majorité des auteurs s'inspirent de [Geman et Geman'84], en proposant un catalogue exhaustif de configurations possibles, pour le système de voisinage considéré, avec une pénalité (constante) par configuration. Des valeurs proches de zéro sont attribuées aux configurations désirables (discontinuité qui se prolonge dans l'espace ou absence de discontinuité), tandis que les configurations indésirables (discontinuités isolées, croisements, discontinuités parallèles ...) reçoivent des valeurs élevées. Le principal inconvénient de cette approche est la nécessité de fixer expérimentalement un grand nombre de paramètres (constantes associées à chaque configuration). Pour s'en affranchir, on peut diviser le catalogue en deux catégories : configurations acceptables (pénalité nulle) et configurations interdites (pénalité infinie) [Heitz et Bouthémy'93]. Il est important de noter que la mise en interaction de processus de ligne, de façon à introduire un a priori géométrique sur les frontières, modifie sensiblement l'aspect algorithmique de l'optimisation. En effet, les variables interstitielles deviennent alors interdépendantes.

4.3. estimateurs robustes

Rappelons que le rôle d'un processus de ligne est de suspendre localement la validité du modèle continu (lisse) du champ principal. Cette notion peut être généralisée à toutes les situations où il est nécessaire de suspendre localement la validité d'un modèle statistique quelconque. Dans le cadre de l'estimation bayésienne, il peut s'agir, par exemple, du modèle gaussien des observations (16) utilisant la mesure d'erreur ε définie par la formule (13). Ce modèle est inadapté hors des zones de validité de l'équation de contrainte de mouvement (1) et, en particulier, pour les mouvements rapides, dans les zones de variations abruptes de la luminosité et dans les zones d'occultation.

La généralisation de la notion de processus de ligne, appelée en anglais *outlier process* (ce qui pourrait être traduit comme « processus de mesures aberrantes »), a été étudiée en détail dans [Black'92b], [Black et Rangarajan'93 et 94]. Les *outlier processes* constituent un trait d'union entre les processus de ligne et les estimateurs issus de la statistique robuste. En effet, le principe de la statistique robuste consiste à admettre d'avance que certaines données ne correspondront pas au modèle utilisé et à essayer de réduire leur influence sur le résultat. Une bonne revue des estimateurs robustes se trouve dans [Meer et al.'91]. Les auteurs retracent d'abord l'historique des techniques d'estimation robuste, s'appuyant sur un classement en M -, R - et L -estimateurs [Huber'81]. Puis ils se concentrent sur un estimateur plus récent, moindre médiane des carrés (ang. *least median of squares*), utilisé en estimation du mouvement par [Cloutier et al.'94]. Les M -estimateurs sont utilisés nettement plus souvent que les autres techniques d'estimation robuste. Ils minimisent une somme $\sum \phi(r_i)$, où la fonction $\phi(r_i)$ est symétrique, définie positive et l'erreur résiduelle r_i est définie comme la différence entre la valeur estimée courante et la donnée pour le i -ème point de mesure. Par abus de langage, nous utiliserons par la suite le terme estimateur pour désigner la fonction associée. Les autres familles d'estimateurs robustes, qui ne seront pas détaillées davantage, utilisent les statistiques d'ordre des erreurs résiduelles. Ceci demande de classer les r_i (ou leurs carrés) dans l'ordre croissant de leurs valeurs, par exemple : $r^{(0)} \leq r^{(1)} \leq \dots \leq r^{(m)} \leq \dots \leq r^{(2N-1)} r^{(2N)}$. On appelle m -ième statistique d'ordre, notée $r^{(m)}$ la valeur de l'élément de rang m , dans l'ensemble des r_i (le rang m n'est pas à confondre avec l'indice i du point de mesure) [Bolon'95]. L'utilisation des statistiques d'ordre consiste à choisir les éléments d'un rang particulier (par exemple, on choisit souvent la valeur médiane, c'est-à-dire la statistique d'ordre N) ou à effectuer une combinaison particulière des statistiques de différents ordres.

Pour expliquer le principe des estimateurs robustes en général et des M -estimateurs en particulier, nous allons commencer par un exemple. Supposons que l'on souhaite estimer les paramètres (a, b) d'une loi linéaire $y = ax + b$, à partir d'un ensemble de mesures $\{(x_i, y_i)\}$. Lorsque toutes les mesures sont cohérentes, le nuage des points de mesure épouse bien la forme de la droite recherchée, mais d'éventuelles mesures erronées

(aberrantes) peuvent donner lieu à des points éloignés de la droite. Une estimation au sens des moindres carrés, c'est-à-dire avec $\phi(r_i) = r_i^2$, aboutit alors à une droite passant entre les deux nuages de points (fig. 8), car la fonction quadratique confère une forte influence aux erreurs résiduelles grandes. L'élimination des mesures aberrantes permettrait pourtant d'identifier la droite correctement.

Une manière de le faire consiste à réduire l'influence des grandes valeurs de r_i en utilisant une fonction $\phi(r_i)$ appropriée : quadratique pour les petites valeurs de r_i mais ayant une croissance plus lente à partir d'une certaine valeur de r_i . Cette influence est mesurée par la dérivée de la fonction $\phi(r_i)$. Un des premiers M -estimateurs, appelé minimax, a été proposé par Huber en 1964. Cet estimateur combine deux fonctions : quadratique pour $|r_i| \leq \beta$, et valeur absolue au-delà :

$$\begin{aligned} \phi_{H\beta}(r_i) &= \begin{cases} r_i^2/2\beta + \frac{\beta}{2} & \text{pour } |r_i| \leq \beta \\ |r_i| & \text{pour } |r_i| > \beta \end{cases} \\ \phi'_{H\beta}(r_i) &= \begin{cases} r_i/\beta & \text{pour } |r_i| \leq \beta \\ \text{sign}(r_i) & \text{pour } |r_i| > \beta \end{cases} \end{aligned} \quad (32)$$

L'influence des erreurs résiduelles est constante au-delà de β . Un comportement similaire est obtenu avec la fonction de Green (fig. 9b) :

$$\phi_{G_{pq}}(r_i) = \frac{2p^2}{q^4} \ln \left[\cosh \left(\frac{q^2 r_i}{p^2} \right) \right] \quad (33)$$

où p et q sont des paramètres réels. Pour limiter davantage l'influence des mesures aberrantes, on utilise souvent des estimateurs « redescendants », c'est-à-dire ayant une fonction d'influence qui décroît plus ou moins rapidement au delà d'une certaine valeur de r_i . C'est notamment le cas de la fonction quadratique tronquée (fig. 9a) [Blake et Zissermann'87], introduite dans le domaine de reconstruction d'images, dont la fonction d'influence passe à zéro

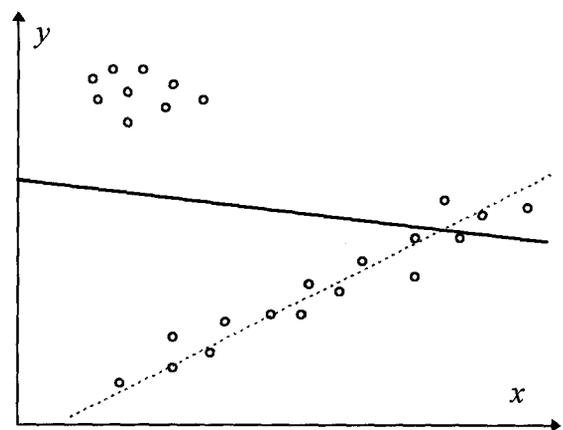


Figure 8. - Estimation, au sens des moindres carrés, des paramètres (a, b) d'une loi linéaire $y = ax + b$ (représentée par une droite pointillée), à partir d'un ensemble de mesures $\{(x_i, y_i)\}$. Le résultat (droite continue) est biaisé par la présence d'un nuage de points de mesure aberrants.

pour $|r_i| \geq \sqrt{\beta/\alpha}$:

$$\begin{aligned} \phi_{BZ_{\alpha\beta}}(r_i) &= \begin{cases} \alpha r_i^2 & \text{si } |r_i| < \sqrt{\beta/\alpha} \\ \beta & \text{sinon} \end{cases} \\ \phi'_{BZ_{\alpha\beta}}(r_i) &= \begin{cases} 2\alpha r_i & \text{si } |r_i| < \sqrt{\beta/\alpha} \\ 0 & \text{sinon} \end{cases} \end{aligned} \quad (34)$$

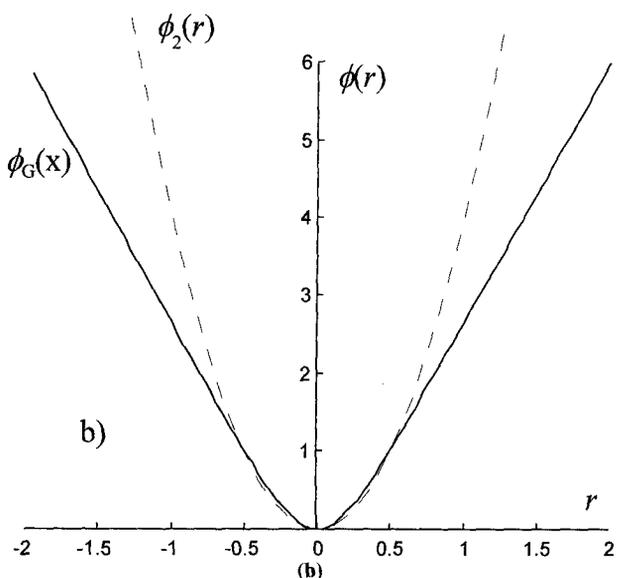
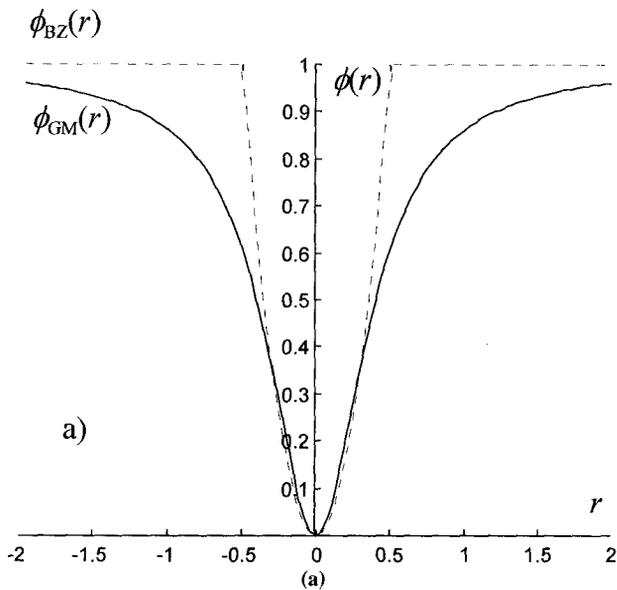


Figure 9. – Exemples de fonctions utilisées comme estimateurs robustes (les coefficients ont été choisis de manière à avoir la partie quadratique commune pour toutes les fonctions et l'asymptote horizontale égale à 1) : (a) estimateurs avec asymptote horizontale, de Blake-Zissermann (34) (en pointillé, $\alpha = 4$, $\beta = 1$) et de Geman-McClure (36) ($p = 0.4$, $q = 1$), (b) estimateur à asymptote oblique de Green (33) avec $1/p^2 = 0.3$ et $q = 1$, comparé à la fonction quadratique $\phi_2(r) = 4r^2$ (en pointillé).

Blake et Zissermann ont montré que l'on aboutit à cet estimateur en partant d'un processus de ligne binaire L . La démarche consiste à minimiser l'énergie des deux champs couplés (principal et dual) d'abord par rapport à la variable L . La même démonstration a été faite pour les *outlier processes* [Black'92b], [Black et Rangarajan'93 et 94]. Le même raisonnement s'applique aussi aux processus de ligne continus et aboutit, sous certaines conditions, à une autre classe d'estimateurs robustes, utilisée en restauration d'images [Geman et Reynolds'92] :

$$\phi_{GR_\delta}(r_i) = \frac{-1}{1 + |r_i|/\delta} \quad \text{avec } \delta > 0 \quad (35)$$

Notons que le lien entre les processus de ligne, *outlier processes* et les estimateurs robustes a également été discuté dans [Black et Anandan'93]. Les estimateurs (34) et (35) appartiennent à une large famille d'estimateurs à asymptote horizontale. De cette famille, nous citerons encore deux fonctions exploitées en estimation du mouvement :

$$\phi_{GM_{pq}}(r_i) = \frac{r_i^2}{p^2 + q^2 r_i^2} \quad \text{avec } (p, q) \in \mathbf{R}^2 \quad (36)$$

de Geman-McClure (fig. 9a), utilisée entre autres par [Barlaud et al.'92], [Blanc-Féraud et al.'93] et [Gaidon et al.'93], ainsi que l'estimateur *biweight* de Tukey :

$$\phi_{Tp}(r_i) = \begin{cases} p^4 r_i^2 / 2 - p^2 r_i^4 / 2 + r_i^6 / 6 & \text{si } |r_i| < p \\ p^6 / 6 & \text{sinon} \end{cases} \quad (37)$$

utilisé par exemple dans [Odohez et Bouthémy'95]. L'utilisation des estimateurs robustes en estimation du mouvement, comme dans d'autres problèmes nécessitant la régularisation, consiste généralement à reformuler de manière robuste, dans (2), le terme d'attache aux données et le terme régularisant (ou au moins l'un des deux). Dans le cadre du formalisme markovien, par exemple, on peut remplacer (22) par :

$$U_d(o/e) = \sum_{s \in S} \phi_d[V_s(e)] \quad \text{et} \quad U_s(e) = \sum_{c \in C} \phi_s[V_c(e)] \quad (38)$$

où ϕ_s et ϕ_d peuvent être deux estimateurs différents. La minimisation est généralement convertie en un problème de moindres carrés pondérés itératifs (cf section 4.5). A chaque itération, le poids énergétiquement optimal, pour r_i fixé, est déterminé par la fonction $w(r_i) = \frac{\phi'(r_i)}{2r_i}$, appelée fonction d'interaction. Outre son utilité pour la résolution numérique du problème de minimisation, la fonction d'interaction est elle-même exploitable. En effet, elle agit comme un *outlier process* implicite. Par conséquent, l'image des valeurs de $w(r_i)$ représente une carte des points aberrants. Lorsque l'estimateur robuste est utilisé dans le terme régularisant, ces points aberrants correspondent généralement à des discontinuités. Lorsque l'estimateur robuste est utilisé dans le terme d'attache aux données, les points aberrants correspondent aux zones situées hors du domaine de validité du modèle des observations,

zones d'occultation par exemple. Dans le cas de l'identification du mouvement dominant [Black et Jepson'94], [Cloutier *et al.*'94], [Odohez et Bouthémy'94 et 95], les points aberrants sont situés dans les régions effectuant des mouvements différents du mouvement dominant.

Notons que l'application des estimateurs robustes dans le domaine d'estimation du mouvement n'est pas limitée au cadre markovien. Un estimateur robuste peut, par exemple, être utilisé lors d'une estimation initiale du mouvement par mise en correspondance de blocs. Plutôt que de maximiser la corrélation entre les blocs, la mise en correspondance consiste souvent à minimiser la somme des différences quadratiques (ang. *sum of squared differences*) des intensités. Là encore, la fonction quadratique peut être remplacée par une fonction ϕ appropriée [Black et Anandan'90b et 91], [Black'92a].

4.4. méthodes variationnelles et diffusion anisotrope

Les méthodes exposées dans cette section relèvent d'une démarche variationnelle, donc résolument continue, dont le formalisme repose sur des équations aux dérivées partielles (EDP) de diffusion. Par un choix convenable de la fonctionnelle à minimiser, on essaie de contrôler l'amplitude et la direction du lissage du champ de vitesses réalisant ainsi une diffusion anisotrope. Nagel a proposé un lissage sélectif du champ de vitesses (cf § 2.2.2), mais celui-ci s'appuie sur les gradients spatio-temporels de la fonction d'intensité [Nagel'87 et 90], alors qu'ici on contrôle le lissage à l'aide des gradients des composantes de la vitesse estimée. Dans un premier temps, nous présentons un formalisme unifié valable pour la majorité des méthodes. Puis nous citons deux méthodes sortant de ce schéma.

Le formalisme utilisé ci-après, dans le cadre de l'estimation du mouvement, a été initialement développé et étudié en amélioration et restauration d'images. Une très bonne synthèse est présentée dans [Deriche et Faugeras'95]. Le problème est posé sous la forme générale suivante : déterminer le couple de fonctions $\mathbf{v} = (u, v)$ qui minimise la fonctionnelle :

$$U(u, v) = \iint \left[(I_x u + I_y v + I_t)^2 + \alpha (\phi(\|\nabla u\|) + \phi(\|\nabla v\|)) \right] dx dy \quad (39)$$

où $\phi(\cdot)$ est une fonction qui permet de contrôler le lissage du champ de mouvement et assure, si possible, une solution continue \mathbf{v} unique et stable. Les équations d'Euler-Lagrange résultantes font apparaître un terme de divergence qui peut se décomposer en deux composantes, l'une dans la direction du gradient notée η , l'autre dans la direction orthogonale au gradient notée ξ :

$$\operatorname{div} \left(\frac{\phi'(\|\nabla w\|)}{\|\nabla w\|} \nabla w \right) = \frac{\phi'(\|\nabla w\|)}{\|\nabla w\|} w_{\xi\xi} + \phi''(\|\nabla w\|) w_{\eta\eta} \quad (40)$$

où $w_{\eta\eta}$ et $w_{\xi\xi}$ sont les dérivées secondes directionnelles, dans les directions η et ξ respectivement. La notation w représente indifféremment u ou v , dans les équations qui sont identiques pour les deux composantes de \mathbf{v} . L'anisotropie de la diffusion provient de la différence des termes de pondération $\frac{\phi'(\|\nabla w\|)}{\|\nabla w\|}$ et $\phi''(\|\nabla w\|)$. Si les deux termes de pondération de l'équation sont égaux alors la propagation est isotrope. On cherche donc une fonction ϕ telle que le lissage soit favorisé dans les régions de mouvement homogène et limité là où il y a discontinuité de mouvement. Pour obtenir un lissage isotrope dans les régions homogènes, la condition s'écrit :

$$\lim_{\|\nabla w\| \rightarrow 0} \frac{\phi'(\|\nabla w\|)}{\|\nabla w\|} = \lim_{\|\nabla w\| \rightarrow 0} \phi''(\|\nabla w\|) = \phi''(0) > 0 \quad (41)$$

Pour éviter le lissage dans les zones de discontinuité, il faut annuler la pondération de $\phi''(\|\nabla w\|)$ lorsque $\|\nabla w\| \rightarrow \infty$. Ces deux conditions ne pouvant être réunies simultanément, on impose les conditions suivantes pour diminuer la diffusion plus rapidement dans la direction du gradient que dans la direction orthogonale au gradient :

$$\begin{aligned} \lim_{\|\nabla w\| \rightarrow \infty} \frac{\phi'(\|\nabla w\|)}{\|\nabla w\|} &= \lim_{\|\nabla w\| \rightarrow \infty} \phi''(\|\nabla w\|) = 0 \\ \lim_{\|\nabla w\| \rightarrow \infty} \frac{\phi''(\|\nabla w\|)}{\frac{\phi'(\|\nabla w\|)}{\|\nabla w\|}} &= 0 \end{aligned} \quad (42)$$

L'existence et l'unicité de la solution avec une fonction ϕ respectant ces conditions a été démontrée [Aubert et Lazaroia'95]. Plusieurs fonctions ϕ peuvent être proposées. [Deriche *et al.*'95] ont utilisé entre autres la fonction $\phi(\|\nabla w\|) = \sqrt{1 + (\|\nabla w\|/k)^2} - 1$. Les auteurs proposent d'effectuer la minimisation de l'énergie U par l'un des deux schémas numériques suivants : l'un résout les équations non-linéaires d'Euler-Lagrange par la méthode itérative de Gauss-Seidel, l'autre considère une nouvelle énergie convexe dont les équations linéaires résultantes sont facilement résolues (régularisation semi-quadratique). On peut constater ici une ressemblance avec les techniques de régression robuste (cf section 4.3).

On peut remarquer que, dans le cas où $\phi(\|\nabla w\|) = \|\nabla w\|^2$, les deux termes de pondération sont égaux, on réalise donc un lissage isotrope (Horn et Schunck). En revanche, si l'on prend $\phi(\|\nabla w\|) = \|\nabla w\|$, la pondération selon la direction du gradient est annulée, on réalise alors une diffusion anisotrope. Cette fonction est exploitée dans [Cohen'93] et [Kumar *et al.*'95]. Les équations d'Euler-Lagrange résultantes sont non-linéaires. La méthode des différences finies ne donnant pas de très bons résultats dans ce cas, [Cohen'93] utilise une méthode d'éléments finis. [Kumar *et al.*'95] utilise une autre démarche introduisant la courbure des courbes de niveau des fonctions recherchées. En effet, le laplacien intervenant dans les équations d'Euler-Lagrange du problème classique avec régularisation quadratique

Estimation du flot optique

est remplacé ici par le terme en divergence suivant :

$$\begin{aligned} & \frac{\partial}{\partial x} \left(\frac{w_x}{\sqrt{w_x^2 + w_y^2}} \right) + \frac{\partial}{\partial y} \left(\frac{w_y}{\sqrt{w_x^2 + w_y^2}} \right) \\ &= \frac{w_x^2 w_{yy} - 2w_x w_y w_{xy} + w_y^2 w_{xx}}{(w_x^2 + w_y^2)^{\frac{3}{2}}} = \operatorname{div} \left(\frac{\nabla w}{\|\nabla w\|} \right) = K_w \end{aligned} \quad (43)$$

où K_w représente la courbure des ensembles de lignes de niveau de la fonction w . Etant donnée l'expression des équations d'évolution associées, l'estimation du flot optique peut donc être vue comme un problème d'évolution de courbes. Une méthode numérique particulière, proposée dans [Osher et Sethian'88], est employée pour la recherche de la solution.

La diffusion anisotrope est également utilisée dans [Proesmans *et al.*'94], mais le schéma est un peu particulier. De l'équation de contrainte de mouvement (1), on déduit une expression approchée de la dérivée temporelle :

$$\begin{aligned} I_t(\tilde{u}_+, \tilde{v}_+) &= -I_x u - I_y v \\ &+ \frac{I(x, y, t) - I(x - \tilde{u}_+ dt, y - \tilde{v}_+ dt, t - dt)}{dt} \end{aligned} \quad (44)$$

où $(\tilde{u}_+, \tilde{v}_+)$ est l'estimation courante de la vitesse apparente (u, v) en (x, y) , calculée entre l'image (t) et l'image $(t - dt)$. Elle peut aussi être calculée entre l'image $(t - dt)$ et l'image (t) , la notation sera alors $(\tilde{u}_-, \tilde{v}_-)$. Les approximations $(\tilde{u}_+, \tilde{v}_+)$ et $(\tilde{u}_-, \tilde{v}_-)$ sont utilisées dans les équations d'évolution. Pour un point donné, les deux approximations sont similaires dans des régions homogènes au sens du mouvement, elles sont discordantes en présence d'une discontinuité ou d'une occultation. Ceci est utilisé pour la détection des discontinuités par une mesure de la concordance des deux mesures de vitesse. De plus, pour éviter une sur-détection des discontinuités, on applique un premier processus de diffusion sur les mesures de concordance. Pour contrôler le lissage du champ des vitesses, un terme de diffusion anisotrope de type Perona-Malik [Perona et Malik'90] est ajouté dans les équations différentielles de contrainte du flot optique. On obtient 6 équations de diffusion non-linéaires couplées : pour chacune des deux approximations de la vitesse $(\tilde{u}_+, \tilde{v}_+)$ et $(\tilde{u}_-, \tilde{v}_-)$ on a deux équations pour la vitesse (une par composante) et une pour la mesure de concordance. Par exemple, l'expression pour la composante u de la vitesse, avec l'approximation $(\tilde{u}_+, \tilde{v}_+)$, est :

$$\frac{\partial u}{\partial t} = \operatorname{div}(\gamma(\|\nabla u\|)\nabla u) - \lambda^2 I_x [I_x u + I_y v + I_t(\tilde{u}_+, \tilde{v}_+)] \quad (45)$$

où la constante λ a un rôle semblable à la constante de régularisation α dans (2) et $\gamma(\|\nabla u\|)$ est une des fonctions proposées par Perona et Malik, paire, strictement décroissante sur \mathbf{R}^+ . On a la même expression pour la composante v .

Citons encore la méthode décrite dans [Nesi'93], faisant appel à la notion de Γ -convergence. Cette approche demande l'introduction d'une variable auxiliaire z prenant des valeurs comprises entre 0 (en présence d'une discontinuité) et 1 (dans une zone homogène). La fonctionnelle à minimiser prend la forme suivante :

$$\begin{aligned} U(u, v, z) &= \iint [(I_x u + I_y v + I_t)^2 + \alpha z^2 (u_x^2 + u_y^2 + v_x^2 + v_y^2) \\ &+ \beta \left(\frac{\|\nabla z\|^2}{k} + \frac{k(1-z)^2}{4} \right)] dx dy \end{aligned} \quad (46)$$

avec $\alpha, \beta > 0$. On reconnaît, dans le premier terme, la contrainte de mouvement classique. Le deuxième terme correspond au terme de régularisation de Horn et Schunck, mais il est contrôlé par la variable z . Le troisième terme est une pénalité qui dépend de la présence de discontinuités. Des résultats théoriques montrent que $\lim_{k \rightarrow \infty} U(u, v, z) = U(u, v)$. Pour résoudre les trois équations d'Euler-Lagrange, associées aux inconnues u, v, z , on utilise une discrétisation par différences finies et la méthode de Jacobi.

4.5. techniques de minimisation de critères non-convexes

Comme il a été dit dans l'introduction, la régularisation du problème de l'estimation de mouvement nécessite généralement la minimisation d'une fonctionnelle d'énergie (2). C'est notamment le cas des méthodes qui, utilisant la modélisation par champs de Markov, cherchent le maximum de la probabilité a posteriori (MAP). Les algorithmes déterministes classiques (gradient, Newton, Gauss-Newton ...) ont été conçus pour la minimisation de critères convexes. La méthode classique utilisée très fréquemment dans l'approche MAP est l'algorithme déterministe ICM (ang. *Iterated Conditional Modes*). Malgré le risque de rester « piégé » dans un minimum local, certains auteurs utilisent cet algorithme même lorsque la modélisation des discontinuités (utilisant un processus de ligne par exemple) rend non convexe le critère à minimiser [Heitz et Bouthémy'89], [Stiller'93]. D'autres, comme [Murray et Buxton'87], [Geman et Reynolds'92], [Konrad et Dubois'92], préfèrent l'algorithme stochastique de recuit simulé (ang. *simulated annealing*) inspiré des processus thermodynamiques. L'algorithme de recuit simulé converge en théorie vers un minimum global à condition d'observer une décroissance en loi logarithmique du paramètre de température. On peut noter ici une variante intéressante de cet algorithme, appelée ISM (*Incremental Stochastic Minimization*) et adaptée aux longues séquences d'images [Black et Anandan'90a et 91], [Black'92a et b]. Contrairement au recuit simulé, où l'on cherche la convergence pour des données figées (un seul couple d'images), ISM travaille sur des données qui évoluent. En effet, au lieu de faire de nombreuses itérations pour un couple d'images (diminuant progressivement la température jusqu'à un niveau très bas) avant de passer à l'image suivante, cet algorithme consiste à faire évoluer

la température au cours de la séquence, sans avoir atteint la convergence à un instant donné. Cette démarche vise évidemment à réduire le temps de calcul total, tout en améliorant la solution au fil du temps. Il existe également des algorithmes déterministes, à convergence plus rapide que les algorithmes stochastiques, conçus avec le souci de s'approcher au mieux du minimum global. Des exemples de tels algorithmes ont déjà été cités plus haut. Ici, nous faisons une synthèse rapide des approches utilisées le plus fréquemment. Tous les algorithmes cités sont itératifs et consistent à minimiser une série de critères, convexes ou non. Malgré ce dénominateur commun, on peut proposer un classement en quatre catégories : multi-résolution, non-convexité graduelle, recuit en champ moyen et relaxation successive.

4.5.1. approche multi-résolution

L'approche multi-résolution est généralement utilisée dans un schéma descendant (ang. *coarse to fine*) où le résultat obtenu à un niveau de résolution grossier (en haut de la pyramide d'images) sert d'initialisation pour les calculs à un niveau plus fin. Les calculs, et en particulier les critères à minimiser, sont les mêmes à chaque niveau de résolution [Barlaud *et al.*'92], [Blanc-Féraud *et al.*'93], [Gaidon *et al.*'93], [Chang *et al.*'94]. L'utilisation d'un tel schéma multi-résolution sert avant tout à accélérer la convergence par rapport aux calculs effectués directement au niveau le plus fin. Dans le cas de l'estimation du mouvement, cette approche permet aussi d'augmenter l'amplitude des mouvements qui peuvent être estimés. Pratiquement n'importe quel algorithme de minimisation utilisé en mono-résolution peut être appliqué à tous les niveaux d'un tel schéma multi-résolution.

La méthode proposée dans [Heitz *et al.*'93] et [Perez et Heitz'93] se distingue de l'approche multi-résolution classique par deux aspects. D'une part, elle ne demande pas la construction d'une pyramide d'images (ou, plus exactement, d'observations), mais seulement d'une pyramide d'étiquettes. D'autre part, les modèles markoviens, et par conséquent les critères à minimiser, ne sont pas identiques à tous les niveaux de résolution, mais sont déduits du critère de départ correspondant à la résolution la plus fine. La minimisation se fait cependant dans le schéma classique descendant, à l'aide de l'algorithme ICM.

L'approche multi-résolution n'offre pas de mécanisme spécifique pour la minimisation de critères non convexes. Néanmoins, la motivation de son utilisation est liée à l'initialisation des calculs à un niveau fin par le champ obtenu au niveau de résolution plus grossier. En effet, on espère que cette initialisation sera suffisamment proche du minimum global, pour éviter le piège des minima locaux.

4.5.2. non-convexité graduelle

Une idée semblable a motivé le développement de l'approche non-convexité graduelle dont le nom vient de l'algorithme GNC (ang. *Graduated Non-Convexity*) [Blake et Zissermann'87],

[Black et Anandan'93], [Black et Rangarajan'93 et 94], [Black et Jepson'94]. Il s'agit d'initialiser le champ en utilisant un critère convexe, sans tenir compte des discontinuités. L'estimation est ensuite améliorée par l'introduction graduelle des discontinuités à travers une série de critères de moins en moins convexes. Soit par exemple l'estimateur lorentzien [Black et Anandan'93], [Black'94], défini par :

$$\phi_{\delta}(r_i) = \ln \left[1 + \frac{1}{2} (r_i/\delta)^2 \right] \quad (47)$$

Le paramètre δ détermine non seulement la « largeur » de la courbe représentative de cette fonction (fig. 10), mais surtout la valeur de l'erreur au delà de laquelle les données sont considérées comme non conformes au modèle (ang. *outliers*). En effet, la fonction d'influence $\phi'(r_i)$ associée à l'estimateur défini par (47) décroît au-delà $r_i = \sqrt{2}\delta$. Ainsi, il suffit de fixer initialement une valeur élevée de δ , pour obtenir un critère « tolérant » qui ne rejette aucune donnée, donc ne tient pas compte des discontinuités. Après cette initialisation, les calculs sont réitérés avec δ de plus en plus petit. Le nombre d'itérations nécessaires pour obtenir une solution stable, ainsi que la capacité à approcher l'optimum global, dépendent de la loi de décroissance du paramètre δ . Dans [Li'95] par exemple, δ diminue d'une itération à l'autre selon la loi $\delta_{i+1} = \kappa\delta_i$ où la constante κ est expérimentalement fixée dans la fourchette $0.9 \leq \kappa \leq 0.99$.

Il existe une version incrémentale de l'algorithme GNC [Black'92b]. Tout comme dans le cas de l'algorithme ISM (version incrémentale du Recuit Simulé), au lieu de chercher la convergence à un instant donné, on cherche à améliorer l'estimation au fil du temps. Pour ce faire, la valeur du paramètre δ est progressivement diminuée au cours de la séquence traitée.

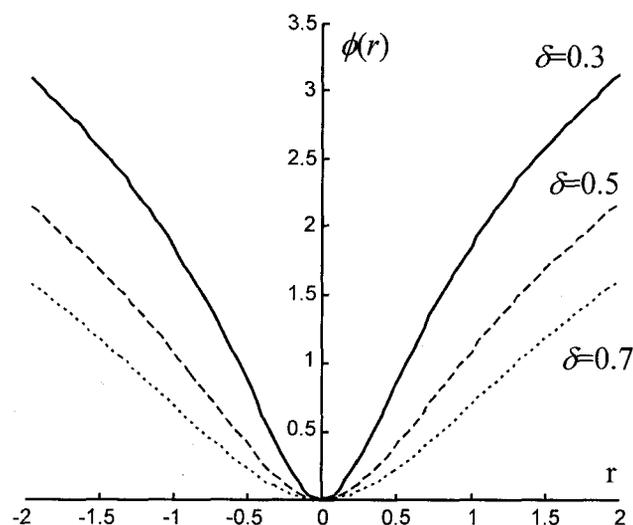


Figure 10. – Représentation graphique de la fonction d'erreur définie par (47) pour trois valeurs différentes du paramètre δ .

4.5.3. recuit en champ moyen

La théorie du champ moyen, comme le recuit simulé, s'inspire de phénomènes thermodynamiques [Zerubia et Chellapa'93]. En dépit de ce point commun, l'algorithme de recuit en champ moyen (ang. *mean field annealing*) est déterministe : au lieu de simuler stochastiquement une distribution de Gibbs, il cherche à l'approcher analytiquement. On montre que le recuit en champ moyen a beaucoup de ressemblances avec le GNC [Bilbro *et al.*'92] et qu'il est particulièrement adapté aux champs markoviens où chaque site interagit avec un petit voisinage [Zhang et Hanauer'93 et 95], [Tian et Shah'94], [Bernier'95]. La dernière des références citées concerne une méthode dans laquelle le recuit en champ moyen est considéré comme équivalent à un réseau d'automates continus synchrones. Pour approcher l'optimum, l'auteur applique son algorithme en quatre passes, en modifiant à chaque passe certains paramètres qui déterminent la prise en compte des discontinuités. La première passe s'effectue sans tenir compte des discontinuités qui sont introduites ultérieurement. Le nombre de passes étant très limité, l'introduction des discontinuités n'est cependant pas aussi progressive que dans l'algorithme GNC.

4.5.4. relaxation successive

Le terme «relaxation successive» englobe une large classe de méthodes itératives consistant à figer à tour de rôle certaines variables pendant que l'on optimise par rapport à d'autres variables. Nous allons aborder ici le domaine privilégié d'application de la relaxation successive, concernant l'utilisation d'estimateurs robustes. Comme il a été dit plus haut, la minimisation avec un M -estimateur est généralement convertie en un problème de moindres carrés pondérés itératifs. L'idée consiste à introduire une variable auxiliaire, transformant le critère non convexe initial en un critère semi-quadratique, convexe vis-à-vis de la variable principale et de la variable auxiliaire. Les deux variables sont figées alternativement. A chaque itération, la variable auxiliaire détermine le poids énergétiquement optimal, pour r_i fixé. La valeur de cette variable est donnée par la fonction d'interaction $w(r_i) = \frac{\phi'(r_i)}{2r_i}$. Après avoir minimisé l'énergie, pour les poids fixés, on recalcule les poids avec les nouvelles valeurs de r_i .

L'existence d'un tel critère dual a été démontrée par [Geman et Reynolds'92]. Cette démonstration repose sur des hypothèses assez contraignantes quant à la forme de la fonction $\phi(r_i)$, exigeant notamment une asymptote horizontale. Ensuite, il a été démontré que ce n'est pas la forme de la fonction $\phi(r_i)$ elle-même, mais la forme de la fonction d'interaction $w(r_i)$ qui détermine la capacité à préserver les discontinuités. Les mêmes conclusions ont été obtenues dans [Li'95] par analyse de l'équation d'Euler. Ceci a permis de poser des conditions moins restrictives que celles de Geman et Reynolds et de démontrer l'existence de fonctions qui permettent d'obtenir un critère dual semi-quadratique, sans avoir une asymptote horizontale [Charbonnier *et al.*'93b], [Charbonnier'94], [Aubert *et al.*'94a]. D'autre part, il a été démontré que

l'algorithme de minimisation du critère dual, consistant à figer à tour de rôle les variables principale et auxiliaire, converge. Le minimum atteint peut éventuellement être local, mais il existe au moins deux familles de fonctions : les fonctions de Green (33) et $\phi(r_i) = \gamma|r_i| - \gamma^2 \ln \left(1 + \frac{|r_i|}{\gamma} \right)$ pour $\gamma > 0$ qui remplissent toutes les conditions quant à la forme de la fonction d'interaction et qui permettent d'obtenir un critère convexe. La relaxation successive garantit alors la convergence vers le minimum global.

5. occultations

Parlant des zones d'occultation, on distingue souvent la zone découverte (celle qui n'était pas visible dans l'image précédente et qui apparaît dans l'image courante) et la zone recouverte (visible dans l'image précédente et cachée dans l'image courante). Cette distinction peut être illustrée dans le cas des méthodes d'estimation par mise en correspondance. Soit une méthode consistant à effectuer une recherche « en avant », c'est-à-dire à prendre les blocs de l'image précédente et à chercher leurs correspondants dans l'image courante. On ne peut pas estimer de cette manière le mouvement des motifs appartenant à une zone recouverte, car ces motifs n'ont pas de correspondant dans l'image courante. Théoriquement, aucun vecteur déplacement ne devrait donc « sortir » d'une telle zone (fig. 11). Inversement, aucun vecteur déplacement ne devrait « arriver » dans une zone découverte, car les motifs appartenant à une zone découverte dans l'image courante n'ont pas d'équivalent dans l'image précédente. Les rôles des deux zones s'inversent lorsque la recherche s'effectue « en arrière ». Dans les deux zones se produit cependant la même chose : l'hypothèse de conservation n'est pas respectée.

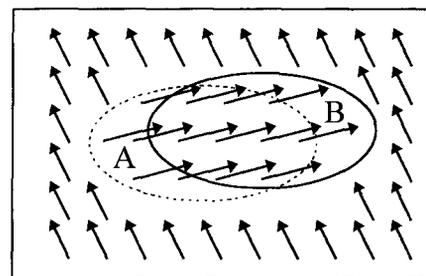


Figure 11. – Champ de vecteurs obtenu par mise en correspondance avec recherche « en avant » : aucun vecteur ne « pointe » vers la zone découverte (A) et aucun vecteur ne « sort » de la zone recouverte (B).

Cette violation de l'hypothèse de conservation met évidemment en cause la validité de l'équation de contrainte de mouvement (1) sur laquelle s'appuient les techniques différentielles, ce qui peut causer des erreurs bien plus importantes que la discontinuité seule [Heitz et Bouthémy '93]. Ceci peut être illustré par un exemple très simple unidimensionnel (fig. 12c). L'équation de contrainte

de mouvement décrit la relation entre le gradient spatial de l'image courante (ici la pente du trait épais), la dérivée temporelle I_t (distance verticale entre les courbes tracées en trait épais d'une part et discontinu d'autre part) et la vitesse (distance horizontale entre les mêmes courbes). La vitesse calculée à l'aide de cette relation est correcte tant que la pente est constante (c'est le cas de \tilde{v}_1). La relation entre la vitesse et les dérivées spatio-temporelles perd son sens dans la zone d'occultation (cas de \tilde{v}_2). En effet, la validité de l'équation (1) est limitée au cas où une facette approximativement plane de la fonction de luminance, « glisse » sur elle-même. Le problème est difficile, car la détection des zones

d'occultation (assimilable à une segmentation) et l'estimation du mouvement sont imbriquées.

On peut citer quatre techniques visant à localiser les zones d'occultation. La solution employée le plus fréquemment consiste à seuiller l'erreur de compensation. Après la mise en évidence des zones où l'erreur de compensation est élevée, on peut limiter voire inhiber la contribution des estimations initiales correspondantes, dans le processus de régularisation. Le cadre qui s'y prête bien est encore une fois le formalisme markovien. On attribue tout simplement un potentiel constant élevé au site correspondant [Stiller'93], [Stiller et Hürtgen'93], [Chang *et al.*'93]. La dernière des méthodes citées utilise à la fois l'information du mouvement et l'information des intensités, effectuant conjointement l'estimation du mouvement et la segmentation. Les poids relatifs des termes « mouvement » et « intensité » dépendent de la fiabilité du mouvement estimé. Ainsi, dans les zones d'occultation, un poids nettement plus important est donné aux intensités. Ceci semble logique, puisque le fond découvert a de fortes chances d'être une continuation du fond adjacent visible précédemment. Lorsque la segmentation selon les niveaux de gris permet d'inclure la zone découverte dans une région dont le mouvement a pu être correctement estimé, on peut envisager de propager l'information de mouvement dans la zone concernée. Notons cependant, que la détection des zones imprédictibles (donc, en principe, découvertes) ne peut résoudre le problème des fausses estimations dans les zones recouvertes. Cette difficulté peut être contournée lorsque la prédiction est bidirectionnelle (par exemple, pour les images B, dans le contexte du codage MPEG), ce qui nécessite l'utilisation d'au moins trois images à la fois : précédente, courante et suivante [Bonnaud et Labit'95].

Dans le cas de la mise en correspondance, on peut détecter les zones recouvertes et découvertes en comparant localement le nombre de vecteurs « entrants » (ang. *inflow*) et « sortants » (ang. *outflow*) [Black et Anandan'90a et 91], [Black'92a]. Le rapport de ces nombres est normalement proche de 1, sauf dans les zones d'occultation. Les auteurs ne se contentent pas de détecter ces zones. Pour éviter de propager, lors du lissage, les mesures erronées obtenues dans ces zones, ils leur attribuent un coefficient de confiance faible. La minimisation étant faite à l'aide de l'algorithme incrémental ISM, ce manque de confiance se traduit par une température localement élevée.

Notons que le même mécanisme, consistant à réinitialiser localement la température, a été plus récemment proposé en association avec un *outlier process* (Black'94). Les zones d'occultation et, plus généralement, toutes les zones où l'information de mouvement n'est pas fiable, correspondent alors aux points aberrants (*outliers*) détectés. On peut citer aussi [Zhang et Hanauer'93 et 95] où trois champs couplés sont estimés dans un même processus. En plus d'un champ de vitesses et d'un champ de discontinuités (processus de ligne), déjà évoqués dans les pages précédentes, il y a un champ binaire correspondant aux zones d'occultation. Les auteurs l'appellent « champ de segmentation », mais en réalité ce n'est rien d'autre qu'un *outlier process*.

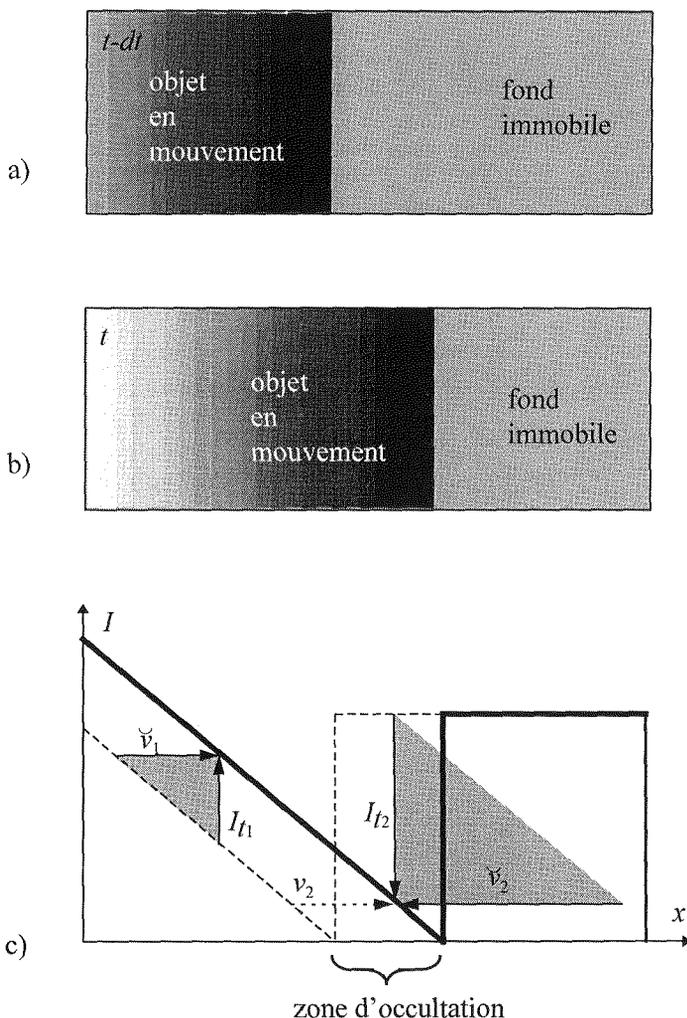


Figure 12. – Fausse estimation de mouvement dans une zone occultée : (a) et (b) images représentant une facette dont le niveau de gris varie linéairement selon l'axe x , qui se déplace vers la droite devant un arrière-plan clair, immobile (c) profils des niveaux de gris de ces images (les traits discontinu et épais correspondant aux instants $t - dt$ et t respectivement). La vitesse estimée pour le point situé dans la zone occultée (\tilde{v}_2) diffère sensiblement de la vitesse réelle (v_2). Les triangles grisés ont été rajoutés pour mettre en évidence la construction géométrique correspondant à l'équation de contrainte de mouvement.

Une autre solution encore vise à vérifier si le point considéré reste au sein d'une même facette linéaire de la fonction de luminance, ce qui revient à vérifier la conservation des dérivées spatiales d'ordre un de cette fonction. La méthode proposée dans [Heitz et Bouthémy'89 et 93] fait appel à un test du rapport de vraisemblance d'hypothèses. Ce test repose sur la modélisation linéaire de la fonction de luminance, dans le voisinage B_x du point $\mathbf{x} = (x, y)$ considéré :

$$I(x + dx, y + dy, t) = I(x, y, t) + a_t dx + b_t dy + n_t \quad (48)$$

$$I(x + dx, y + dy, t + dt) = I(x, y, t + dt) + a_{t+dt} dx + b_{t+dt} dy + n_{t+dt} \quad (49)$$

où n_t et n_{t+dt} sont des termes de bruit, supposés indépendants, gaussiens, centrés, ayant des variances identiques. Les deux hypothèses concurrentes sont alors :

$$H_0 : \{a_t = a_{t+dt} \text{ et } b_t = b_{t+dt}\} \text{ dans } B_x,$$

$$H_1 : \{a_t \neq a_{t+dt} \text{ ou } b_t \neq b_{t+dt}\} \text{ dans } B_x,$$

Selon la première hypothèse, le point considéré appartient à une même facette dans les deux images, tandis que la seconde hypothèse traduit la situation où le point \mathbf{x} , dans l'image courante, n'appartient plus à la même facette que dans l'image précédente. Considérant que l'information du mouvement fournie par un point est utilisable tant que ce point reste sur une même facette, on conserve les observations liées aux points qui vérifient l'hypothèse H_0 et on élimine les observations liées aux points vérifiant l'hypothèse H_1 . Pour ce faire, on introduit, dans (22), une variable binaire $\xi(s)$ qui valide la contribution de chaque site s à l'énergie d'attache aux données U_d . On a alors des potentiels associés à U_d de la forme :

$$V_s = \xi(s) [\nabla I(s) \cdot \mathbf{v}(s) + I_t(s)]^2 \quad (50)$$

Pour palier le manque d'information de mouvement dans les zones marquées par $\xi(s) = 0$, donc surtout dans les zones d'occultation, les auteurs propagent l'information de mouvement obtenue par identification des éléments de contours en mouvement (cf § 2.2.1). Ceci est assuré par une autre fonction de potentiel, que nous n'allons pas détailler, utilisant un processus de ligne trinaire (cf § 4.2.1). Le signe du processus de ligne indique de quel côté du contour doit se faire la propagation.

6. conclusions

Cette revue des techniques d'estimation du mouvement en présence de discontinuités permet de dégager deux tendances. La première met l'accent sur la recherche de zones de continuité

(régions homogènes du point de vue du mouvement) et fait appel à des modèles paramétriques de mouvement. La deuxième, au contraire, se focalise sur la recherche des ruptures locales de continuité et met en jeu des détecteurs de discontinuités, explicites ou implicites. Dans les deux approches, on constate un succès certain des estimateurs robustes qui permettent d'estimer le mouvement dominant (localement), en évitant de biaiser le résultat de cette estimation par l'influence des points voisins effectuant des mouvements différents.

La méthodologie généralement choisie pour régulariser le problème d'estimation du mouvement s'appuie sur l'estimation bayésienne au sens du maximum de probabilité a posteriori (MAP). Ce cadre méthodologique permet d'introduire un a priori sur la structure du champ à estimer, par le biais de la modélisation markovienne. Une approche similaire, mais fondée sur des équations de diffusion aux dérivées partielles, est apparue récemment. Dans les deux cas, on aboutit à la minimisation de critères généralement non convexes. L'effort de nombreux chercheurs porte aujourd'hui sur les algorithmes de minimisation à la fois rapides et capables d'approcher au mieux le minimum global. Le point commun de la majorité des méthodes de minimisation est l'utilisation d'une série de critères, soit tous convexes (méthodes où les variables sont figées à tour de rôle), soit échelonnés entre un critère convexe et le critère original non convexe (non-convexité graduelle). Ce schéma permet l'utilisation d'algorithmes déterministes, pour minimiser chaque critère, tout en réduisant le risque de rester piégé dans un minimum local. Il est appliqué seul ou conjointement à un schéma multi-résolution.

Par opposition aux grandes tendances évoquées ci-dessus, on peut constater un nombre relativement faible de méthodes qui prennent en compte la continuité temporelle, les occultations et la fiabilité des estimations. Ces trois problèmes sont liés. En effet, l'exploitation de la continuité temporelle peut permettre à la fois d'améliorer la fiabilité des estimations là où le contenu de la scène évolue lentement et de localiser les zones découvertes ou recouvertes, donc celles où le contenu de la scène change brusquement donnant des estimations non fiables. Une mesure de confiance peut à son tour permettre de propager uniquement les estimations fiables. Il peut s'agir de la propagation aussi bien au sens spatial (lissage) qu'au sens temporel (prédiction / compensation). La prise en compte de la continuité temporelle, des occultations et de la fiabilité des estimations semble mériter d'être encore explorée.

Remerciements

Ce travail a été effectué dans le cadre du projet DRED : « 3D / discontinuités », financé par le GDR TdSI. Les auteurs remercient Isabelle E. MAGNIN, Bruno NEYRAN et Denis FRIBOULET et surtout les experts de la revue *Traitement du Signal*, pour leurs remarques précieuses concernant autant la forme que le contenu de cet article.

BIBLIOGRAPHIE

- [Aisbett'89] J. Aisbett, « Optical Flow with an Intensity-Weighted Smoothing », *IEEE Trans. PAMI*, 11 : 5, 1989, pp. 512-522.
- [Aubert et al.'94a] G. Aubert, M. Barlaud, L. Blanc-Féraud, P. Charbonnier, « Deterministic Edge-Preserving Regularization in Computed Imaging », techn. report n°94-01, I3S Univ. Nice - Sophia Antipolis, 1994, soumis à *IEEE Trans. Image Proc.*
- [Aubert et al.'94b] G. Aubert, M. Barlaud, L. Blanc-Féraud, P. Charbonnier, « A Deterministic Algorithm for Edge-Preserving Computed Imaging Using Legendre Transform », *Proc. 12th IAPR Int. Conf. Pattern Recogn.*, Jerusalem, Vol. III, Oct. 1994, pp. 188-191.
- [Aubert et Lazaroaia'95] G. Aubert, L. Lazaroaia, « A Variational Method in Image Recovery », rapport URA CNRS #168, Laboratoire J-A Dieudonné, Université de Nice Sophia Antipolis, N°TR 423, 1995.
- [Barlaud et al.'92] M. Barlaud, L. Blanc-Féraud, J.M. Collin, « Motion Estimation Involving Discontinuities in a Multiresolution Scheme », *SPIE Proc. Visual Comm. & Image Proc.*, Boston, Vol. 1818, Nov. 1992, pp. 526-541.
- [Barron et al.'94] J.L. Barron, D.J. Fleet, S.S. Beauchemin, « Performance of Optical Flow Techniques », *Int. J. Comp. Vision*, 12 : 1, 1994, pp. 43-77.
- [Benois et al.'93] J. Benois, L. Wu, D. Barba, « Segmentation spatio-temporelle d'images animées basée sur le fusionnement de régions homogènes spatialement », *14e Colloque GRETSI*, Juan les Pins, 1993, pp. 867-870.
- [Bernier'95] O. Bernier, « Analyse du flot optique dans les séquences d'images par réseau d'automates », *15e Colloque GRETSI*, Juan les Pins, 1995, pp. 833-836.
- [Bilbro et al.'92] G.L. Bilbro, W.E. Snyder, S.J. Garnier, J.W. Gault, « Mean Field Annealing : a Formalism for Constructing GNC-like Algorithms », *IEEE Trans. Neural Networks*, 3 : 1, Jan. 1992, pp. 131-138.
- [Black et Anandan'90a] M.J. Black, P. Anandan, « A Model for the Detection of Motion over Time », research report n°YALEU/DCS/RR-822, Yale Univ. (USA), 1990, et *Proc. Int. Conf. Comp. Vision*, Osaka, Dec. 1990, pp. 33-37.
- [Black et Anandan'90b] M.J. Black, P. Anandan, « Constraints for the Early Detection of Discontinuity from Motion », research report n°YALEU/DCS/RR-789, Yale Univ. (USA), et *Nat. Conf. Artif. Intell. AAAI*, Boston (USA), 1990, pp. 1060-1066.
- [Black et Anandan'91] M.J. Black, P. Anandan, « Robust Dynamic Motion Estimation over Time », *IEEE Comp. Soc. Conf. Comp. Vision & Pattern Recogn.*, Lahaina-Maui (Hawaii), 1991, pp. 296-302.
- [Black'92a] M.J. Black, « Combining Intensity and Motion for Incremental Segmentation and Tracking Over Long Image Sequences », research report n°YALEU/DCS/RR-873, Yale Univ. (USA), 1991 et *2nd Europ. Conf. Comp. Vision*, May 1992, pp. 485-493.
- [Black'92b] M.J. Black, *Robust Incremental Optical Flow*, Ph.D. dissert., Yale Univ. (USA), 1992, 257 pages.
- [Black et Anandan'93] M.J. Black, P. Anandan, « A Framework for the Robust Estimation of Optical Flow », *4th Int. Conf. Comp. Vision, Berlin (Ger.)*, 1993, pp. 231-236.
- [Black et Jepson'94] M.J. Black, A.D. Jepson, « Estimating Multiple Independent Motions in Segmented Images using Parametric Models with Local Deformations », *IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, Austin TX (USA), nov. 1994, 8 pages.
- [Black et Rangarajan'93] M.J. Black, A. Rangarajan, « On Line Processes, Outlier Rejection and Robust Statistics », techn. report n°SPL-93-072, XEROX Palo Alto Research Center, 1993, à paraître dans *Int. J. Comp. Vision*.
- [Black'94] M.J. Black, « Recursive Non-Linear Estimation of Discontinuous Flow Fields », *Europ. Conf. Comp. Vision*, Stockholm (Sweden), Springer Verlag, May 1994, 8 pages.
- [Black et Rangarajan'94] M.J. Black, A. Rangarajan, « The Outlier Process : Unifying Line Processes and Robust Statistics », *Proc. Comp. Vision Pattern Recogn.*, Seattle WA (USA), Jun. 1994, pp. 15-22.
- [Blake et Zisserman'87] A. Blake, A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge (USA), London (GB), 1987.
- [Blanc-Féraud et al.'93] L. Blanc-Féraud, M. Barlaud, T. Gaidon, « Motion Estimation Involving Discontinuities in a Multiresolution Scheme », *Opt. Eng.*, 32 : 7, 1993, pp. 1475-1482.
- [Bolon'95] P. Bolon, « Filtres d'ordre », dans *Analyse d'images : filtrage et segmentation* (J.P. Coqueret, S. Philipp, Eds), Masson, Paris, Milan, Barcelona, 1995, pp. 79-85.
- [Bonnaud et Labit'95] L. Bonnaud, C. Labit, « Codage interpolatif de séquences d'images utilisant un suivi temporel de segmentation spatio-temporelle », *15e Colloque GRETSI*, Juan les Pins, 1995, pp. 821-824.
- [Bouthémy'87] P. Bouthémy, « Estimation et structuration d'indices spatio-temporels pour l'analyse du mouvement dans une séquence d'images », *Traitement du Signal*, 4 : 3, 1987, pp. 239-257.
- [Bouthémy'88] P. Bouthémy, « Modèles et méthodes pour l'analyse du mouvement dans une séquence d'images », *2e Atelier Scientifique Traitement d'Images : du Pixel à l'Interprétation*, Aussois, 1988, XXV-1 à XXV-19.
- [Bouthémy'89] P. Bouthémy, « A Maximum Likelihood Framework for Determining Moving Edges », *IEEE Trans. PAMI*, 11 : 5, May 1989, pp. 499-511.
- [Bouthémy'95] P. Bouthémy, « Interactions entre segmentation et mesure dans l'analyse du mouvement 2D : orientations récentes et perspectives », dans les actes de *l'Ecole Thématique du GDR Traitement du Signal et Images*, Daudon, 3-5 avril 1995.
- [Bouthémy et François'93] P. Bouthémy, E. François, « Motion Segmentation and Qualitative Dynamic Scene Analysis from an Image Sequence », *Int. J. Comp. Vision*, 10 : 2, 1993, pp. 157-182.
- [Brofferio et Mercalli'89] S.C. Brofferio, G. Mercalli, « Motion Estimation Using Pointwise Multigrid Measurements », *12e Colloque GRETSI*, Juan les Pins 1989, pp. 609-612.
- [Chang et al.'93] M.M. Chang, A.M. Tekalp, M.I. Sezan, « Motion Field Segmentation Using an Adaptive MAP Criterion », *ICASSP*, Minneapolis (USA), 1993, V-33 à V-36.
- [Chang et al.'94] M.M. Chang, M.I. Sezan, A.M. Tekalp, « An Algorithm for Simultaneous Motion Estimation and Scene Segmentation », *ICASSP*, Adelaide (Australia), 1994, V-221 à V-224.
- [Charbonnier et al.'93a] P. Charbonnier, L. Blanc-Féraud, M. Barlaud, « An Adaptive Reconstruction Method Involving Discontinuities », *ICASSP*, Minneapolis (USA), 1993, V-491 à V-494.
- [Charbonnier et al.'93b] P. Charbonnier, L. Blanc-Féraud, M. Barlaud, « ARTUR : An Adaptive Deterministic Relaxation Algorithm for Edge-Preserving Tomographic Reconstruction », techn. report n°93-76, I3S Univ. Nice - Sophia Antipolis, 1993.
- [Charbonnier'94] P. Charbonnier, *Reconstruction d'image : régularisation avec prise en compte des discontinuités*, thèse Univ. Nice - Sophia Antipolis, 1994, 212 pages.
- [Charbonnier et al.'94] P. Charbonnier, L. Blanc-Féraud, G. Aubert, M. Barlaud, « Two Deterministic Half-Quadratic Regularization Algorithms for Computed Imaging », *Int. Conf. Image Proc.*, Austin TX (USA), 3, Nov. 1994, pp. 168-171.
- [Chen et al.'95] W. Chen, G.B. Giannakis, N. Nandhakumar, « A Harmonic Retrieval Framework for Discontinuous Motion Estimation », *Int. Conf. Image Proc.*, Washington DC (USA), 1, 1995, pp. 219-222.
- [Chou et Chen'93] W.S. Chou, Y.C. Chen, « Estimation of the Velocity Field of Two-Dimensional Deformable Motion », *Pattern Recogn.*, 26 : 2, 1993, pp. 351-364.
- [Chu et Delp'89] C.H. Chu, E.J. Delp, « Estimating Displacement Vectors from an Image Sequence », *J. Opt. Soc. Am. A*, 6 : 6, 1989, pp. 871-878.
- [Cloutier et al.'94] L. Cloutier, A. Mitiche, P. Bouthémy, « Segmentation and Estimation of Image Motion by a Robust Method », *Int. Conf. Image Proc.*, Austin TX (USA), 2, 1994, pp. 805-809.
- [Cohen'93] I. Cohen, « Non-linear variational method for optical flow computation », *Proc. 8th SCIA*, Vol. 1, 1993, pp. 523-530.
- [Deriche et Faugeras'95] R. Deriche, O. Faugeras, « Les EDP en traitement des images et vision par ordinateur », rapport INRIA N°TR 2697, novembre 1995.

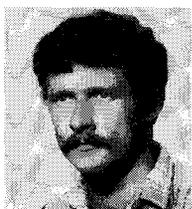
- [Deriche *et al.*'95] R. Deriche, P. Kornprobst, G. Auber, « Optical-flow estimation while preserving its discontinuities : a variational approach », *Proc. 2nd ACCV*, Singapore, 2, Dec. 1995, pp. 71–80.
- [Devlamincq *et al.*'95] V. Devlamincq, L. Lelaurin, A. Yacoubi, J.P. Dubus, « Estimation of Discontinuous Optical Flow Using a Neural Network for Echocardiographic Images », *Proc. SPIE Medical Imaging*, San Diego CA (USA), Vol. 2434, Feb. 1995.
- [Diehl'91] N. Diehl, « Object-Oriented Motion Estimation and Segmentation in Image Sequences », *Signal Proc. : Image Comm.*, 3, 1991, pp. 23–56.
- [Duc *et al.*'95] B. Duc, P. Schroeter, J. Bigün, « Motion Estimation and Segmentation by Fuzzy Clustering », *Int. Conf. Image Proc.*, Washington DC (USA), 3, 1995, pp. 472–475.
- [Etoh et Shirai'93] M. Etoh, Y. Shirai, « Segmentation and 2D Motion Estimation by Region Fragments », *Int. Conf. Comp. Vision*, Berlin (Ger.), 1993, pp. 192–199.
- [François et Bouthémy'90] E. François, P. Bouthémy, « Derivation of Qualitative Information in Motion Analysis », *Image & Vision Computing*, 8 : 4, 1990, pp. 279–288.
- [François et Bouthémy'91] E. François, P. Bouthémy, « Temporally-Linked Motion-Based Segmentation in a Long Image Sequence », *7th Scand. Conf. Image Anal.*, Aarlborg (DK), 1991, pp. 1134–1141.
- [François'91] E. François, *Interprétation qualitative du mouvement à partir d'une séquence d'images*, thèse Univ. Rennes, 1991, 278 pages.
- [Fua'91] P. Fua, « A Parallel Stereo Algorithm that Produces Dense Depth Maps and Preserves Image Features », Rapport de Recherche n°1369, INRIA Sophia Antipolis, 1991.
- [Gaidon *et al.*'93] T. Gaidon, M. Barlaud, P. Mathieu, « Compression de séquences d'images avec estimation de mouvement et quantification vectorielle algébrique », *14e Colloque GRETSI*, Juan les Pins, 1993, pp. 871–874.
- [Garcia-Garduno et Labit'93] V. Garcia-Garduno, C. Labit, « Suivi de segmentation spatio-temporelle pour la compression de séquences animées », *14e Colloque GRETSI*, Juan les Pins, 1993, pp. 847–850.
- [Geman et Geman'84] S. Geman, D. Geman, « Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images », *IEEE Trans. PAMI*, 6 : 6, 1984, pp. 721–741.
- [Geman et Reynolds'92] D. Geman, G. Reynolds, « Constrained Restoration and the Recovery of Discontinuities », *IEEE Trans. PAMI*, 14 : 3, 1992, pp. 367–383.
- [Germain et Baudois'93a] F. Germain, D. Baudois, « Segmentation Based on Motion : An Incremental Approach Using Kalman Temporal Filter », *Proc. SPIE Math. Methods for Medical Imaging*, San Diego CA (USA), Vol. 2035, 1993, pp. 120–131.
- [Germain et Baudois'93b] F. Germain, D. Baudois, « Approche adaptative de l'estimation du mouvement dans une séquence d'images par filtrage de Kalman temporel », *14e Colloque GRETSI*, Juan les Pins, 1993, pp. 875–878.
- [Germain et Skordas'94] F. Germain, T. Skordas, « An Image Motion Estimation Technique Based on a Combined Statistical Test and Spatio-Temporal Generalized Likelihood Ratio Approach », *3rd ECCV*, Stockholm, May 1994, pp. 152–157.
- [Germain'95] F. Germain, *Estimation du mouvement dans une séquence d'images, en contexte linéaire optimal*, thèse INP de Grenoble, CEPHAG, 1995, 161 pages.
- [Ghosal et Vanek'96] S. Ghosal, P. Vanek, « A fast scalable algorithm for discontinuous optical flow estimation », *IEEE Trans. Pattern Anal. Machine Intell.*, 18 : 2, 1996, pp. 181–194.
- [Heeger'87] D.J. Heeger, « Optical Flow from Spatio-Temporal Filters », *Proc. Int. Conf. Comp. Vision*, 1987, pp. 181–190.
- [Heitz et Bouthémy'89] F. Heitz, P. Bouthémy, « Estimation et segmentation du mouvement : approche bayésienne et modélisation markovienne des occlusions », *7e Congrès AFCET / INRIA Recon. Formes Intell. Artif.*, Paris, 1989, pp. 1359–1368.
- [Heitz et Bouthémy'93] F. Heitz, P. Bouthémy, « Multimodal Estimation of Discontinuous Optical Flow Using Markov Random Fields », *IEEE Trans. PAMI*, 15 : 12, 1993, pp. 1217–1233.
- [Heitz *et al.*'93] F. Heitz, P. Perez, P. Bouthémy, « Multiscale Minimization of Global Energy Functions in Some Visual Recovery Problems », *Comp. Vision Graph. Image Proc. : Image Understanding*, 59 : 1, Jan. 1994, pp. 125–134.
- [Horn et Schunck'81] B.K.P. Horn, B. Schunck, « Determining Optical Flow », *Artif. Intell.*, 17, 1981, pp. 185–203.
- [Huber'81] P.J. Huber, *Robust Statistics*, Wiley : New York, 1981, pp. 185–203.
- [Jacobson et Wechsler'87] L. Jacobson, H. Wechsler, « Derivation of Optical Flow Using Spatio-Temporal Frequency Approach », *Comp. Vision Graphics Image Proc.*, 38, 1987, pp. 29–65.
- [Jepson et Black'93] A.D. Jepson, M.J. Black, « Mixture Models for Optical Flow Computation », techn. report n°RBCV-TR-93-44, Toronto Univ. (Canada) et *Conf. Vision Pattern Recogn.*, Jun. 1993, pp. 760–761 (version résumée) et dans : *Partitioning Data Sets with Applications to Psychology, Vision and Target Tracking*, DIMACS Workshop, (AMS Publ.), Providence (USA), Vol. 19, 1995, pp. 271–286.
- [Konrad et Dubois'92] J. Konrad, E. Dubois, « Bayesian Estimation of Motion Vector Fields », *IEEE Trans. PAMI*, 14 : 9, 1992, pp. 910–927.
- [Kottke et Sun'94] D.P. Kottke, Y. Sun, « Motion Estimation via Cluster Matching », *IEEE Trans. PAMI*, 16 : 11, 1994, pp. 1128–1133.
- [Kumar *et al.*'95] A. Kumar, A. Tannenbaum, G. Balazs, « Optical flow : a curve evolution approach », *Int. Conf. Image Proc.*, Washington DC (USA), 3, 1995, pp. 17–20.
- [Lai *et al.*'94] J. Lai, J. Gauch, J. Crisman, « Computing Optical Flow in Color Image Sequences », *Innov. Technol. Biol. Med.*, 15 : 3, 1994, pp. 375–390.
- [LeQuang *et al.*'95] D. LeQuang, A. Zaccarin, S. Caron, « Object-Oriented Coding Using Successive Motion Field Segmentation and Estimation », *Int. Conf. Image Proc.*, Washington DC (USA), 1, 1995, pp. 207–210.
- [Li'95] S.Z. Li, « On Discontinuity-Adaptive Smoothness Priors in Computer Vision », *IEEE Trans. PAMI*, 17 : 6, 1995, pp. 576–586.
- [Meer *et al.*'91] P. Meer, D. Mintz, A. Rosenfeld, D.Y. Kim, « Robust Regression Methods for Computer Vision : A Review », *Int. J. Comp. Vision*, 6 : 1, 1991, pp. 59–70.
- [Mitiche *et al.*'87] A. Mitiche, Y.F. Wang, J.K. Aggarwal, « Experiments in Computing Optical Flow with the Gradient-Based Multiconstraint Method », *Pattern Recogn.*, 20 : 2, 1987, pp. 173–179.
- [Mitiche et Bouthémy'96] A. Mitiche, P. Bouthémy, « Computation and Analysis of Image Motion : a Synopsis of Current Problems and Methods », *Int. J. Comp. Vision*, 19 : 1, 1996, pp. 29–55.
- [Méménin et Perez'95] E. Méménin, P. Perez, « Champs de Markov multirésolution et algorithme multigrille pour l'estimation de mouvement », *15e Colloque GRETSI*, Juan les Pins, 1995, pp. 813–816.
- [Murray et Williams'86] D.W. Murray, N.S. Williams, « Detecting the Image Boundaries between Optical Flow Fields from Several Moving Planar Facets », *Pattern Recogn. Letters*, 4, 1986, pp. 87–92.
- [Murray et Buxton'87] D.W. Murray, B.F. Buxton, « Scene Segmentation from Visual Motion Using Global Optimization », *IEEE Trans. PAMI*, 9 : 2, 1987, pp. 220–228.
- [Nagel'87] H.H. Nagel, « On the Estimation of Optical Flow : Relations between Different Approaches and Some New Results », *Artif. Intell.*, 33, 1987, pp. 299–324.
- [Nagel'90] H.H. Nagel, « Extending the 'Oriented Smoothness Constraint' into the Temporal Domain and the Estimation of Derivatives of Optical Flow », *1st Europ. Conf. Comp. Vision, Antibes (France)*, 1990, pp. 139–148.
- [Nesi'93] P. Nesi, « Variational approach to optical flow estimation managing discontinuities », *Image and Vision Comp.*, 11 : 7, Sept. 1993, pp. 419–439.
- [Odobez et Bouthémy'94] J.M. Odobez, P. Bouthémy, « Detection of Multiple Moving Objects Using Multiscale MRF with Camera Motion Compensation », *Int. Conf. Image Proc.*, Austin TX (USA), 2, 1994, pp. 257–261.

- [Odohez et Bouthémy'95] J.M. Odohez, P. Bouthemy, « Estimation robuste multiéchelle de modèles paramétrés de mouvement sur des scènes complexes », *Traitement du Signal*, 12 : 2, 1995, pp. 113–128.
- [Osher et Sethian'88] S.J. Osher, J.A. Sethian, « Fronts propagating with curvature dependent speed : algorithms based on Hamiltonian–Jacobi formulations », *Journal of Computational Physics*, 79, 1988, pp. 12–49.
- [Pérez et Heitz'93] P. Perez, F. Heitz, « Une approche multiéchelle à l'analyse d'images par champs markoviens », *Traitement du Signal*, 9 : 6, 1993, pp. 459–472.
- [Perona et Malik'90] P. Perona, J. Malik, « Scale–Space and Edge Detection Using Anisotropic Diffusion », *IEEE Trans. PAMI*, 12 : 7, 1990, pp. 629–639.
- [Proesmans et al.'94] M. Proesmans, L. Van Gool, E. Pauwels, A. Oosterlinck, « Determination of optical flow and its discontinuities using non–linear diffusion », *Europ. Conf. Comp. Vision*, Springer Verlag, 1994, pp. 295–304.
- [Sanson'93] H. Sanson, « Estimation de mouvement basée sur des régions pour le codage de séquences télévisuelles », *14e Colloque GRETSI*, Juan les Pins, 1993, pp. 859–862.
- [Sanson'95a] H. Sanson, « Vers une méthodologie pour l'identification paramétrique robuste du mouvement de régions par optimisation non–linéaire », *15e Colloque GRETSI*, Juan les Pins, 1995, pp. 817–820.
- [Sanson'95b] H. Sanson, « Toward a Robust Parametric Identification of Motion on Regions of Arbitrary Shape by Non–Linear Optimization », *Int. Conf. Image Proc.*, Washington DC (USA), 1, 1995, pp. 203–206.
- [Schnörr'91] C. Schnörr, « Determining Optical Flow for Irregular Domains by Minimizing Quadratic Functionals of a Certain Class », *Int. J. Comp. Vision*, 6 : 1, 1991, pp. 25–38.
- [Schnörr'92] C. Schnörr, « Computation of Discontinuous Optical Flow by Domain Decomposition and Shape Optimization », *Int. J. Comp. Vision*, 8 : 2, 1992, pp. 153–165.
- [Schunck'89] B.G. Schunck, « Image Flow Segmentation and Estimation by Constraint Line Clustering », *IEEE Trans. PAMI*, 11 : 2, 1989, pp. 1010–1027.
- [Singh'91] A. Singh, « Optic Flow Computation : a Unified Perspective », *IEEE Comp. Soc. Press : Los Alamitos CA (USA)*, 1991.
- [Stiller'93] C. Stiller, « A Statistical Image Model for Motion Estimation », *ICASSP*, Minneapolis (USA), 1993, V–193 à V–196.
- [Stiller'94] C. Stiller, « Object–Oriented Video Coding Employing Dense Motion Fields », *ICASSP*, Adelaide (Australia), 1994, V–273 à V–276.
- [Stiller et Hürtgen'93] C. Stiller, B. Hürtgen, « Combined Displacement Estimation and Segmentation in Image Sequences », *Proc. Fibre Optic Networks Video Compression*, Berlin (Germany), April 1993.
- [Terzopoulos'86] D. Terzopoulos, « Regularization of Inverse Visual Problems Involving Discontinuities », *IEEE Trans. PAMI*, 8 : 4, 1986, pp. 413–423.
- [Thiaville et al.'91] L. Thiaville, P. Puget, J.–J. Niez, « Un réseau de neurones pour le calcul du mouvement », *Neuro–Nîmes'91 4th Int. Conf. Neural Networks & Their Appl.*, 1991, pp. 59–70.
- [Tian et Shah'94] T.Y. Tian, M. Shah, « Motion Segmentation and Estimation », *Int. Conf. Image Proc.*, Austin TX (USA), 2, 1994, pp. 785–789.
- [Thompson et al.'85] W.B. Thompson, K.M. Mutch, V.A. Berzins, « Dynamic Occlusion Analysis in Optical Flow Fields », *IEEE Trans. PAMI*, 7 : 4, 1985, pp. 374–383.
- [Tistarelli'95] M. Tistarelli, « Computation of Coherent Optical Flow by Using Multiple Constraints », *Int. Conf. Comp. Vision*, Cambridge (MA), 1995, pp. 263–268.
- [Wohn et al.'83] K. Wohn, L.S. Davis, P. Thrift, « Motion Estimation Based on Multiple Local Constraints and Nonlinear Smoothing », *Pattern Recogn.*, 16 : 6, 1983, pp. 563–570.
- [Xiong et Graffigne'94] W. Xiong, C. Graffigne, « A Hierarchical Method for Detection of Moving Objects », *Int. Conf. Image Proc.*, Austin TX (USA), 2, 1994, pp. 795–799.
- [Zerubia et Chellapa'93] J. Zerubia, R. Chellapa, « Mean Field annealing using compound Gauss–Markov Random fields for edge detection and image estimation », *IEEE Transactions on Neural Networks*, 4 : 4, July 1993, pp. 703–709.
- [Zetsche et Barth'91] C. Zetsche, E. Barth, « Direct Detection of Flow Discontinuities by 3D Curvature Operators », *Pattern Recogn. Letters 12*, 1991, pp. 771–779.
- [Zhang et Hanauer'93] J. Zhang, G.G. Hanauer, « The Mean Field Theory for Image Motion Estimation », *ICASSP*, Minneapolis (USA), 1993, V–197 à V–200.
- [Zhang et Hanauer'95] J. Zhang, G.G. Hanauer, « The Application of Mean Field Theory to Image Motion Estimation », *IEEE Trans. Image Proc.*, 4 : 1, 1995, pp. 19–33.
- [Zhang et al.'95] K. Zhang, M. Bober, J. Kittler, « Motion Based Segmentation for Video Coding », *Int. Conf. Image Proc.*, Washington DC (USA), 3, 1995, pp. 476–479.
- [Zheng et Blostein'92] W.J. Zheng, S.D. Blostein, « Adaptive Regularization for Motion Field Estimation », *Signal Proc. VI : Theories and Appl.*, 1992, pp. 1327–1330.

Manuscrit reçu le 20 octobre 1995.

LES AUTEURS

Maciej ORKISZ



Maciej Orkisz est né en 1961 à Cracovie (Pologne). Il est titulaire d'un diplôme d'ingénieur électronicien de la Polytechnique de Varsovie (1985) et d'un doctorat en automatique et traitement du signal de l'Institut National Polytechnique de Grenoble (1990). Depuis 1992, il est maître de conférences à l'Institut National des Sciences Appliquées de Lyon, où il effectue sa recherche au sein du Centre de Recherche et d'Applications en Traitement de l'Image et du Signal (CREATIS). Ses domaines d'intérêt sont avant tout l'estimation du mouvement et la segmentation dynamique, appliquées à l'imagerie bio-médicale.

Patrick CLARYSSE



Patrick Clarysse est titulaire d'un doctorat en automatique de l'Université de Lille, obtenu en 1991. Le sujet de sa thèse concernait le développement d'un système de localisation 3D assistée par ordinateur, pour les besoins de la chirurgie stéréotaxique. Depuis octobre 1992, il travaille au Centre de Recherche et d'Applications en Traitement de l'Image et du Signal (CREATIS), en qualité de chercheur CNRS, département des Sciences de la Vie. Ses principaux domaines d'intérêt sont l'extraction des frontières 3D, analyse locale et globale du mouvement, ainsi que les modèles déformables, appliqués à l'estimation du mouvement du cœur.