

Une approche multiéchelle à l'analyse d'images par champs markoviens

A Multiscale Approach to Image Analysis Using Markov Random Fields



Patrick PEREZ

IRISA/INRIA,
Campus Universitaire de Beaulieu,
35042 Rennes Cedex.

Patrick Pérez est né en 1968. Diplômé de l'École Centrale de Paris, option Mathématiques Appliquées, en 1990, il prépare depuis deux ans un doctorat en traitement du signal à l'IRISA, à Rennes.

Ses domaines d'intérêt actuels sont l'analyse multirésolution de signaux, le traitement de problèmes d'optimisation globale par des méthodes déterministes ou stochastiques, et les modèles markoviens et leur application à l'analyse d'image.



Fabrice HEITZ

IRISA/INRIA,
Campus Universitaire de Beaulieu,
35042 Rennes Cedex.

tion et Modélisation d'Images Séquentielles») à l'IRISA, à Rennes. Diplômé de Télécom Bretagne (1984), il a préparé un doctorat en automatique et traitement du signal à Télécom Paris (1988). Ses premiers travaux ont porté sur l'analyse non destructive d'œuvres d'art par des techniques d'analyse statistique de l'image, en collaboration avec le Laboratoire de Recherches des Musées de France.

Ses domaines d'intérêt actuels couvrent la vision par ordinateur, la modélisation statistique de l'image, l'analyse du mouvement dans des séquences d'images ainsi que la parallélisation des algorithmes de traitement d'images.

Fabrice Heitz est chargé de recherches INRIA. Il mène actuellement ses travaux de recherche dans le projet TEMIS (« Traitement Exploita-

RÉSUMÉ

L'introduction d'une modélisation statistique par champs de Markov a récemment permis des avancées importantes dans nombre de problèmes classiques en analyse d'images. Ces modèles sont généralement associés à des algorithmes d'optimisation globale par relaxation qui restent coûteux en temps de calcul dans certaines applications. Or les techniques multigrilles, par ailleurs classiques en analyse numérique, peuvent conduire à des gains importants sur ce point. Pour l'heure il n'existe cependant pas réellement de support théorique permettant d'associer de façon simple et efficace stratégie multigrille et modélisation markovienne. Les modèles markoviens multirésolutions utilisés pour l'instant sont donc élaborés de manière heuristique, en particulier en ce qui concerne l'ajustement des différents paramètres en fonction de la résolution.

Nous présentons dans cet article une approche multiéchelle de la modélisation markovienne qui est à la fois mathématiquement cohérente et facilement implantable. Nous détaillons son application sur deux exemples d'analyse du mouvement dans une séquence d'images : la détection du mouvement et la mesure du mouvement. Cela permet de mettre en évidence les apports de l'approche : accélération de la convergence et amélioration de l'estimée par rapport aux techniques markoviennes multirésolutions classiques.

MOTS CLÉS

Champs markoviens, analyse d'images, analyse du mouvement, optimisation globale, méthode multigrille, multirésolution et multiéchelle.

ABSTRACT

The use of Markov Random Field (MRF) models within the framework of global bayesian estimation has recently brought new powerful solutions to standard image analysis problems. These models are generally associated with greedy relaxation algorithms. This is the reason why multiresolution methods, well known in Computational Mathematics, are widely used to speed up the convergence rate of these algorithms. But for the moment there is no real mathematical framework which associates in a simple and efficient way multigrad strategies and markovian models : most previous multiresolution markovian models have been defined using various heuristics, especially as far as the adjustment of parameters over scale is concerned.

In this paper we present a new class of globally consistent multiscale

MRF models. The models we consider here are both mathematically consistent and computationally tractable and are related to a multiscale exploration of the set of solutions. We detail the application of these new models to two basic issues in motion analysis from an image sequence : motion detection and 2D-motion estimation. We show the advantages of the new approach : it allows the relaxation schemes to converge faster than those associated with standard multiresolution approaches, and toward better estimates (i.e. estimates of lower energy).

KEY WORDS

Markov Random Fields (MRF), image analysis, motion analysis, global estimation, multigrad, multiresolution and multiscale methods.

1. Introduction

Connus et étudiés par les statisticiens depuis plus de quinze ans [3], les champs de Markov ont fait récemment une apparition remarquable dans le domaine du traitement d'images [14]. D'abord appliqués à des problèmes de restauration [4, 8, 14], ils sont désormais utilisés avec succès dans des domaines aussi variés que l'extraction de contours [13], la segmentation [9, 13], la stéréovision [2], la reconstruction tomographique [11], ou la reconstruction de surface [9, 23].

Le domaine qui nous préoccupe plus particulièrement est celui de l'analyse du mouvement dans des séquences d'images. Riche en problématiques variées, il a donné lieu à de très nombreux travaux ces dix dernières années [1, 6]. Les champs de Markov lui sont appliqués depuis peu pour la détection du mouvement [7], l'estimation de champs 2D de déplacement [17, 20] et la segmentation au sens du mouvement [12, 25] (pour d'autres études utilisant des modèles similaires le lecteur pourra se reporter également à [10, 19, 20, 25, 27]).

Associés à la théorie bayésienne de l'estimation, les champs de Markov offrent un cadre mathématique cohérent et unifié pour l'extraction de primitives, ou étiquettes, à partir d'observations. Le formalisme utilisé permet conjointement :

- l'introduction simple de connaissances a priori sur les primitives
- une modélisation statistique du lien entre les observations et les primitives.

L'estimation des primitives induite in fine par la modélisation markovienne se fait le plus souvent dans un espace de configurations extrêmement vaste et peut donc donner lieu à des optimisations longues et coûteuses. C'est une des raisons qui a motivé l'introduction de techniques multigrilles [28], par ailleurs classiques en analyse numérique [16]. Ces techniques permettent en effet une accélération des algorithmes de relaxation déterministes ou stochastiques utilisés pour l'estimation des primitives. On suppose d'autre part que de telles stratégies fourniront des estimations qualitativement supérieures dans la mesure où l'on intègre l'information de manière graduelle et hiérarchique.

Mais pour l'heure il n'existe pas réellement de support théorique permettant d'associer stratégie multigrille et modélisation markovienne. La méthode généralement retenue consiste à se doter d'une pyramide multirésolution d'observations et de primitives. L'estimation des primitives à une résolution donnée est faite à l'aide d'un modèle de Markov, le même à toutes les résolutions. Seuls les paramètres du modèle peuvent être modifiés d'un niveau à l'autre et ce de façon entièrement heuristique. La coopération entre les différentes résolutions est souvent descendante :

- le champ d'étiquettes est estimé au niveau i ,
- ce champ « projeté » au niveau $i - 1$ (résolution immédiatement supérieure) sert d'initialisation pour l'algorithme d'estimation à ce niveau.

Cette technique a déjà été étudiée dans le cas de la stéréovision [2] et dans celui de l'estimation de champs de déplacements [20, 18]. Si elle donne d'assez bons résultats, elle est cependant dépourvue de fondement théorique quant au choix des modèles markoviens (en général identiques à tous les niveaux) et quant au réglage des paramètres de ces modèles. Dans la pratique les paramètres sont d'ailleurs généralement pris égaux à toutes les résolutions.

S'inspirant de la théorie des « groupes de renormalisation » issue de la physique statistique, Gidas [15] a fourni une approche mathématique cohérente de la modélisation markovienne multirésolution. Malheureusement la formulation obtenue est bien trop complexe pour être mise en œuvre sur les modèles existants hormis celui simple dit modèle d'Ising. Parallèlement Bouman *et al.* [5] présentent une approche simple et séduisante de la question. L'originalité de leur approche réside dans l'estimation d'une structure pyramidale classique de primitives à partir d'un seul niveau d'observations. Mais il subsiste encore dans leur formalisme une grande part d'heuristique au niveau de la modélisation a priori du champ d'étiquettes par des champs de Markov identiques à tous les niveaux (même système de voisinage, mêmes fonctions de potentiel et surtout mêmes valeurs des paramètres de ces fonctions).

Nous présentons dans cet article une approche multiéchelle de la modélisation markovienne qui est à la fois mathématiquement cohérente et facilement implantable. L'algorithme consiste à explorer des espaces de configurations emboîtés, inclus dans l'espace sur lequel le problème d'optimisation traité est originellement posé. On montre que cette technique d'exploration est équivalente à un algorithme multirésolution sur une structure pyramidale de primitives couplée à un seul niveau d'observations, analogue à celle de Bouman *et al.* Dans cette structure particulière, les modèles markoviens aux différents niveaux ne sont plus définis de façon heuristique, mais sont complètement déterminés par le modèle original à la résolution la plus fine, contrairement au modèle markovien multirésolution de Bouman *et al.* Cette approche généralise un travail récent de B. Levy [22] sur les modèles markoviens gaussiens linéaires, aux modèles markoviens non linéaires les plus généraux.

La présentation du cadre mathématique pour la modélisation markovienne multiéchelle fait l'objet de la section 2. La mise en place de ce formalisme nouveau est ensuite détaillée dans le cadre de l'analyse du mouvement (section 3). Il est appliqué d'une part à un modèle markovien de détection du mouvement et d'autre part à un modèle de mesure du champ des vitesses apparentes. Des expérimentations sur scènes réelles démontrent l'efficacité de l'approche par rapport aux techniques multigrilles heuristiques utilisées jusqu'alors.

2. Modèles markoviens multiéchelles

Nous allons avant toute chose rappeler les grandes lignes de la modélisation markovienne et de l'estimation bayésienne globale en traitement d'images. Cela permettra d'introduire les principales notations utilisées dans cet article.

2.1. MODÈLES MARKOVIENS ET ESTIMATION BAYESIENNE GLOBALE

Les observations (informations dont on dispose pour effectuer l'analyse) et les étiquettes (informations que l'on cherche à extraire des observations) sont supposées former des champs de variables aléatoires notés respectivement O (champ d'observations) et E (champ d'étiquettes) et définis respectivement sur des grilles rectangulaires S_1 et S_2 (identiques ou non) :

$$O = \{O_s, s \in S_1\} \quad E = \{E_s, s \in S_2\}$$

où s désigne un site d'une des grilles. On notera $o = \{o_s, s \in S_1\}$ et $e = \{e_s, s \in S_2\}$ une réalisation quelconque de chacun de ces deux champs. Les étiquettes E_s prennent leurs valeurs dans un ensemble discret Λ . L'ensemble de toutes les configurations possibles de E est donc $\Omega = \Lambda^{|S_2|}$ où $|S_2|$ désigne le cardinal de S_2 .

Le critère d'estimation retenu est souvent celui du *Maximum A Posteriori* (MAP) : étant donné une réalisation o du champ d'observations, on cherche la réalisation \hat{e} du champ d'étiquettes la plus probable au sens de la distribution a posteriori. Soit

$$(1) \quad \hat{e} \triangleq \arg \max_{e \in \Omega} p(E = e | O = o).$$

Notons que l'on peut trouver d'autres critères dans la littérature [23].

L'un des intérêts de ce critère est qu'on peut lui appliquer la règle de Bayes, ce qui donne après simplification :

$$(2) \quad \hat{e} = \arg \max_{e \in \Omega} p(o|e) \cdot p(e).$$

Le premier terme de la fonction à maximiser (probabilité d'occurrence des observations conditionnellement aux étiquettes) sera déterminé en modélisant le lien entre les étiquettes et les observations. Le second, probabilité d'occurrence de e , résultera d'une modélisation a priori des propriétés du champ d'étiquettes. Celle-ci se fait simplement en supposant le champ E markovien relativement à un système de voisinage $\mathcal{G} = \{\mathcal{G}_s, s \in S_2\}$ ⁽¹⁾ c'est-à-dire :

$$(i) \quad \forall e \in \Omega, p(E = e) > 0$$

$$(ii) \quad \forall s \in S_2, p(E_s = e_s | E_r = e_r, r \neq s) \\ = p(E_s = e_s | E_r = e_r, r \in \mathcal{G}_s).$$

En vertu du théorème de Hammersley et Clifford [4], E suit une distribution de Gibbs

$$(3) \quad p(E = e) = \pi(e) \triangleq \frac{1}{Z} e^{-U_2(e)},$$

où :

⁽¹⁾ Rappel des propriétés d'un système de voisinage :

1. $\forall s \in S_2, s \notin \mathcal{G}_s$.
2. $\forall \{s, r\} \subset S_2, r \in \mathcal{G}_s \Leftrightarrow s \in \mathcal{G}_r$.

- La fonction $U_2(e)$, dite *fonction d'énergie* se décompose en une somme de fonctions locales V_c , dites *potentiels*, définies sur les cliques $c \in \mathcal{C}$ engendrées par \mathcal{G} ⁽²⁾ :

$$(4) \quad U_2(e) \triangleq \sum_{c \in \mathcal{C}} V_c(e)$$

V_c n'étant en fait fonction que des étiquettes portées par les sites de c .

- Z , constante de normalisation appelée *fonction de partition*, est définie par :

$$(5) \quad Z \triangleq \sum_{e \in \Omega} e^{-U_2(e)}.$$

C'est le choix des potentiels locaux V_c qui va permettre de conférer au champ d'étiquettes des propriétés a priori pertinentes.

Supposons que $p(o|e) > 0, \forall o, \forall e$. En posant $U_1(e, o) = -\ln(p(o|e))$ on voit que \hat{e} est un minimum global de la fonction d'énergie $U(e, o) = U_1(e, o) + U_2(e)$:

$$(6) \quad \hat{e} = \arg \min_{e \in \Omega} U_1(e, o) + U_2(e).$$

La recherche d'un tel minimum est un classique problème d'optimisation. Une solution approchée peut être obtenue soit par des algorithmes déterministes dont l'algorithme ICM (*iterative conditional modes*) est un exemple [4], soit par des algorithmes stochastiques de type « recuit simulé » [14]. Les algorithmes stochastiques convergent théoriquement vers un minimum global de la fonction d'énergie, mais avec une lenteur souvent rédhibitoire pour des espaces de configurations très grands. On leur préfère donc généralement des algorithmes déterministes. Ceux-ci, s'ils convergent beaucoup plus rapidement, présentent cependant l'inconvénient de se piéger dans d'éventuels minima locaux de la fonction d'énergie et risquent donc de converger vers des estimations de qualité inférieure. Pour pallier ce défaut on peut initialiser ces algorithmes avec des configurations supposées proches d'un minimum global. C'est cette idée qui est mise en œuvre dans les techniques multigrilles descendantes.

2.2. MODÈLES MULTIÉCHELLES

Lorsqu'on cherche à traiter par des techniques multigrilles un problème modélisé à l'aide de champs de Markov, la principale difficulté réside dans l'ajustement des différents paramètres en fonction de la résolution. Comme il est montré dans [15], le système de voisinage peut d'autre part varier d'un niveau à l'autre. Dans [15] Gidas a adapté l'approche par groupe de renormalisation issue de la physique statistique pour élaborer un modèle markovien multirésolution cohérent. Cette approche postule essentiellement que l'énergie d'une réalisation du champ de Markov étudié ne change pas avec le niveau de résolution. Malheu-

⁽²⁾ \mathcal{C} est l'ensemble des parties non vides de S_2 réduites à un singleton ou dont les éléments sont voisins deux à deux.

reusement la méthode décrite par Gidas n'est vraiment utilisable que pour des modèles très simples. Bouman *et al.* [5] ont récemment présenté une approche différente basée sur l'estimation d'une pyramide de primitives à partir d'un seul niveau d'observations. Cependant cette modélisation reste heuristique, en particulier en ce qui concerne l'adaptation des paramètres en fonction de la résolution.

L'approche que nous présentons, tout en abordant de façon cohérente le problème de multirésolution markovienne et celui associé, de conservation d'énergie à travers la structure multiéchelle, fournit une modélisation facilement implantable et qui peut être considérée comme une généralisation des travaux de B. Levy [22]. Nous présentons maintenant de façon assez générale ce modèle markovien multiéchelle.

Supposons que les seules observations dont nous disposons pour établir notre modèle multiéchelle soient définies sur une grille S correspondant à l'échelle la plus fine, et que les étiquettes recherchées soient définies sur le même ensemble de sites ⁽³⁾ :

$$E = \{E_s, s \in S\} \quad O = \{O_s, s \in S\}.$$

Supposons d'autre part que l'étiquette E_s en un site s ne soit liée aux observations que par l'intermédiaire de l'observation O_s au même site. Alors l'énergie $U_1(e, o)$ se décompose de la façon suivante :

$$(7) \quad U_1(e, o) = \sum_{s \in S} W_s(e_s, o_s).$$

Cette décomposition est celle qui apparaît effectivement dans nombre des modèles markoviens existants. Quoi qu'il en soit, dans le cas où l'hypothèse précédente n'est pas réalisée, dans certains problèmes de restauration d'images par exemple [14], il est facile d'adapter les calculs qui vont suivre.

Enfin supposons que le système de voisinage \mathcal{G} soit constitué des 8-voisinages ou voisinages d'ordre 2, pour lesquels un site s a 8 voisins. Il existe alors 10 types de cliques, composées de 1, 2, 3 ou 4 sites mutuellement voisins (*fig. 1*). Ce système de voisinage est le plus répandu ; des voisinages d'ordre supérieur sont cependant parfois utilisés. Ce cas pourrait aussi être traité par ce qui va suivre.

L'idée d'une approche multiéchelle est de résoudre le problème (6) dans des espaces de configurations emboîtés $\Omega^n \subset \Omega^{n-1} \subset \dots \subset \Omega^i \subset \dots \subset \Omega^1 \subset \Omega^0 = \Omega$ correspondant à des configurations de plus en plus « fines », dans un sens que nous allons préciser. Soit à estimer la suite de configurations :

$$(8) \quad \hat{e}(i) \triangleq \arg \min_{e \in \Omega^i} U_1(e, o) + U_2(e).$$

Définissons maintenant ces sous-ensembles Ω^i de configurations. Supposons que la grille rectangulaire S ait des

⁽³⁾ La méthode proposée se généralise facilement au cas où les deux grilles S_1 et S_2 sont différentes.

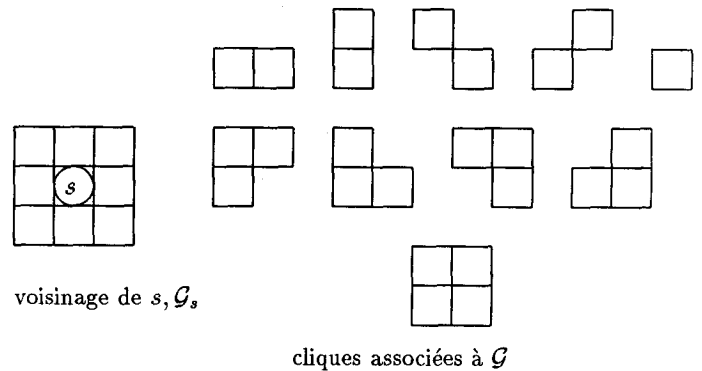


Figure 1. — Voisinage d'ordre 2 et cliques associées.

dimensions de la forme $2^m \times 2^{m'}$, m et $m' \in \mathbb{N}^*$; alors on peut la partitionner en N_i blocs carrés de $2^i \times 2^i$ sites, i vérifiant $i < \inf(m, m')$. Notons B_k^i , $k = 1, \dots, N_i$ ces blocs et B^i la grille rectangulaire qu'ils forment. On a $N_i = 2^{m+m'-2i}$. Le sous-ensemble Ω^i des champs d'étiquettes à l'échelle i est l'ensemble des configurations $e \in \Omega$ constantes sur les blocs B_k^i , $k = 1, \dots, N_i$:

$$(9) \quad e \in \Omega^i \Leftrightarrow \forall k \in \{1, \dots, N_i\}, \exists e_k^i \in \Lambda \mid \forall s \in B_k^i, e_s = e_k^i,$$

e_k^i étant donc l'étiquette commune à tous les sites du bloc B_k^i pour le champ $e \in \Omega^i$ (*fig. 2*). Notons qu'il est tout à fait possible de donner une définition plus complexe à l'ensemble Ω^i , comme par exemple l'ensemble des champs bilinéaires par morceaux sur les blocs B_k^i [26]. Nous nous en tiendrons dans le présent article à la première définition.

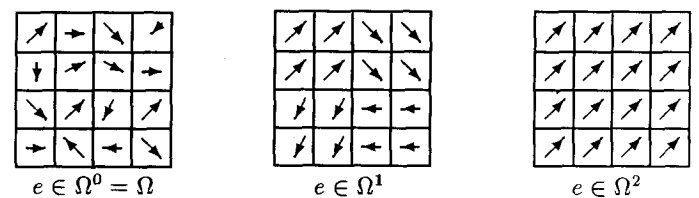


Figure 2. — Exemples de configurations de $\Omega^0 = \Omega$, Ω^1 et Ω^2 (les étiquettes sont ici des vecteurs).

Nous allons maintenant expliciter la fonction d'énergie $U(e, o)$ pour une échelle i donnée. Pour cela on réécrit les fonctions d'énergie (4) et (7) en utilisant la propriété (9). Soit $e \in \Omega^i$:

$$(10) \quad \begin{aligned} U_1(e, o) &= \sum_{s \in S} W_s(e_s, o_s) \\ &= \sum_{B_k^i \in B^i} \sum_{s \in B_k^i} W_s(e_s, o_s) \end{aligned}$$

que l'on peut formuler :

$$(11) \quad U_1(e, o) = \sum_{B_k^i \in B^i} W_k^i(e_k^i, o)$$

avec :

$$(12) \quad W_k^i(e_k^i, o) \triangleq \sum_{s \in B_k^i} W_s(e_k^i, o_s).$$

Étant donnée une clique c de \mathcal{C} , c peut être positionnée de quatre façons différentes par rapport à la grille de blocs B_k^i :

- c est incluse dans un bloc $B_k^i \in B^i$ (fig. 3 a) :

$$c \subset B_k^i.$$

Soit \mathcal{C}_k^i l'ensemble de ces cliques.

- c contient au moins deux sites et est « à cheval » sur deux blocs $\{B_{k_1}^i, B_{k_2}^i\}$ (fig. 3 b) :

$$c \subset \bigcup_{l=1}^{l=2} B_{k_l}^i \quad \text{et} \quad c \cap B_{k_l}^i \neq \emptyset \quad \text{pour} \quad l = 1, 2.$$

Soit \mathcal{C}_{k_1, k_2}^i l'ensemble de ces cliques.

- c contient au moins trois sites et est « à cheval » sur trois blocs $\{B_{k_1}^i, B_{k_2}^i, B_{k_3}^i\}$ (fig. 3 c) :

$$c \subset \bigcup_{l=1}^{l=3} B_{k_l}^i \quad \text{et} \quad c \cap B_{k_l}^i \neq \emptyset \quad \text{pour} \quad l = 1, 2, 3.$$

Soit $\mathcal{C}_{k_1, k_2, k_3}^i$ l'ensemble de ces cliques.

- c contient quatre sites et est « à cheval » sur quatre blocs $\{B_{k_1}^i, B_{k_2}^i, B_{k_3}^i, B_{k_4}^i\}$ (fig. 3 d) :

$$c \subset \bigcup_{l=1}^{l=4} B_{k_l}^i \quad \text{et} \quad c \cap B_{k_l}^i \neq \emptyset \quad \text{pour} \quad l = 1, 2, 3, 4.$$

Soit $\mathcal{C}_{k_1, k_2, k_3, k_4}^i$ l'ensemble de ces cliques.

On voit ainsi apparaître une relation de voisinage sur la grille B^i de blocs, définie de la façon suivante :

$$(13) \quad B_{k_1}^i \text{ et } B_{k_2}^i \text{ voisins} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} B_{k_1}^i = B_{k_2}^i \\ \text{ou} \\ \exists c \in \mathcal{C} : c \cap B_{k_1}^i \neq \emptyset, c \cap B_{k_2}^i \neq \emptyset. \end{cases}$$

Il est immédiat que B^i se trouve ainsi dotée d'un système de 8-voisinage. Soit \mathcal{D}^i l'ensemble des cliques associées.

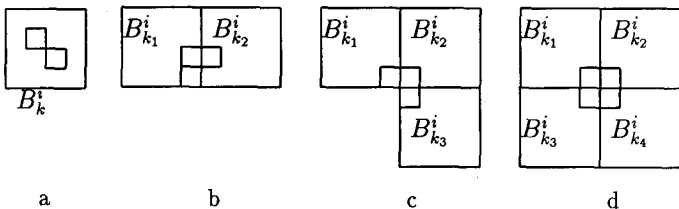


Figure 3. — Exemples de position d'une clique c par rapport aux blocs B_k^i .

L'énergie $U_2(e)$ peut donc se décomposer de la façon suivante :

$$(14) \quad \begin{aligned} U_2(e) &= \sum_{c \in \mathcal{C}} V_c(e) \\ &= \sum_{B_k^i \in B^i} \sum_{c \in \mathcal{C}_k^i} V_c(e) \\ &+ \sum_{\{B_{k_1}^i, B_{k_2}^i\} \in \mathcal{D}^i} \sum_{c \in \mathcal{C}_{k_1, k_2}^i} V_c(e) \\ &+ \sum_{\{B_{k_1}^i, B_{k_2}^i, B_{k_3}^i\} \in \mathcal{D}^i} \sum_{c \in \mathcal{C}_{k_1, k_2, k_3}^i} V_c(e) \\ &+ \sum_{\{B_{k_1}^i, B_{k_2}^i, B_{k_3}^i, B_{k_4}^i\} \in \mathcal{D}^i} \sum_{c \in \mathcal{C}_{k_1, k_2, k_3, k_4}^i} V_c(e) \end{aligned}$$

que l'on peut simplifier :

$$(15) \quad \begin{aligned} U_2(e) &= \sum_{B_k^i \in B^i} V_k^i(e_k^i) \\ &+ \sum_{\{B_{k_1}^i, B_{k_2}^i\} \in \mathcal{D}^i} V_{k_1, k_2}^i(e_{k_1}^i, e_{k_2}^i) \\ &+ \sum_{\{B_{k_1}^i, B_{k_2}^i, B_{k_3}^i\} \in \mathcal{D}^i} V_{k_1, k_2, k_3}^i(e_{k_1}^i, e_{k_2}^i, e_{k_3}^i) \\ &+ \sum_{\{B_{k_1}^i, B_{k_2}^i, B_{k_3}^i, B_{k_4}^i\} \in \mathcal{D}^i} V_{k_1, k_2, k_3, k_4}^i(e_{k_1}^i, e_{k_2}^i, e_{k_3}^i, e_{k_4}^i) \end{aligned}$$

avec :

$$(16) \quad V_k^i(e_k^i) \triangleq \sum_{c \in \mathcal{C}_k^i} V_c(e)$$

$$(17) \quad V_{k_1, k_2}^i(e_{k_1}^i, e_{k_2}^i) \triangleq \sum_{c \in \mathcal{C}_{k_1, k_2}^i} V_c(e)$$

$$(18) \quad V_{k_1, k_2, k_3}^i(e_{k_1}^i, e_{k_2}^i, e_{k_3}^i) \triangleq \sum_{c \in \mathcal{C}_{k_1, k_2, k_3}^i} V_c(e)$$

$$(19) \quad V_{k_1, k_2, k_3, k_4}^i(e_{k_1}^i, e_{k_2}^i, e_{k_3}^i, e_{k_4}^i) \triangleq \sum_{c \in \mathcal{C}_{k_1, k_2, k_3, k_4}^i} V_c(e).$$

Les vraisemblances $W_s(e_s, o_s)$, $s \in S$ et les potentiels $V_c(e)$, $c \in \mathcal{C}$ étant donnés il est alors possible grâce à (11) et (15) d'explicitier simplement l'énergie totale $U(e, o)$ pour les configurations appartenant à l'un des sous-ensembles Ω^i . C'est ce que nous allons faire dans les sections suivantes pour deux exemples de modèles markoviens utilisés en analyse du mouvement. Mais voyons auparavant comment le modèle développé peut être interprété en terme de structuration pyramidale des primitives.

2.3. MODÈLE MULTIÉCHELLE ET MODÈLE PYRAMIDAL

On peut remarquer que le terme $U_2(e)$, $e \in \Omega^i$ donné par (15) se décompose sur les cliques de blocs engendrées par

un système de 8-voisinage. D'autre part chaque bloc B_k^i ne présente qu'une variable à estimer : l'étiquette e_k^i commune aux différents sites qui le composent. Il semble donc naturel pour une échelle donnée, d'assimiler chaque bloc B_k^i à un unique site s d'une nouvelle grille rectangulaire S^i de taille $2^{m-i} \times 2^{m'-i}$ isomorphe à B^i . En particulier S^i est dotée d'un système de 8-voisinage que l'on notera \mathcal{G}^i . Celui-ci engendre un ensemble de cliques \mathcal{C}^i . Chacun de ces sites $s \in S^i$ portera l'étiquette du bloc qu'il représente, et que l'on notera E_s^i . On définit ainsi de nouveaux champs d'étiquettes $E^i = \{E_s^i, s \in S^i\}$ ayant respectivement pour ensemble de configurations $\Gamma^i = \Lambda^{|S^i|}$ isomorphe à Ω^i , avec $\Gamma^0 = \Omega^0 = \Omega$. Notons Φ^i la bijection de Γ^i dans Ω^i (fig. 4) :

$$(20) \quad \begin{aligned} \Phi^i : \Gamma^i &\rightarrow \Omega^i \\ e^i &\mapsto e = \Phi^i(e^i). \end{aligned}$$

On obtient finalement une structure pyramidale des primitives dans le sens où chaque grille S^i est quatre fois plus petite que la « précédente » S^{i-1} et $\Phi^i(e^i)$ est en quelque sorte « l'interpolé » (par simple répétition) de e^i à la résolution la plus fine (fig. 5).

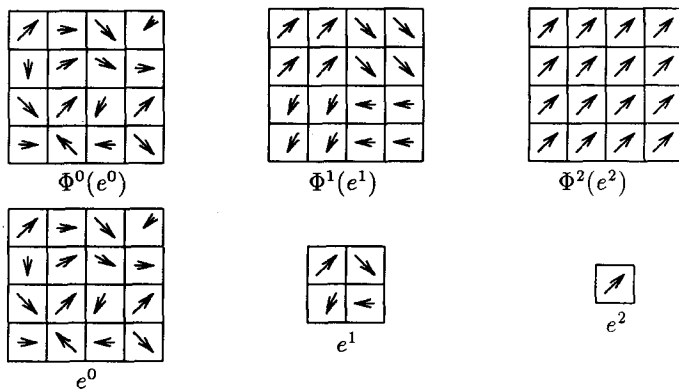


Figure 4. — Exemples de configurations $e^i \in \Gamma^i$ et $\Phi^i(e^i) \in \Omega^i$.

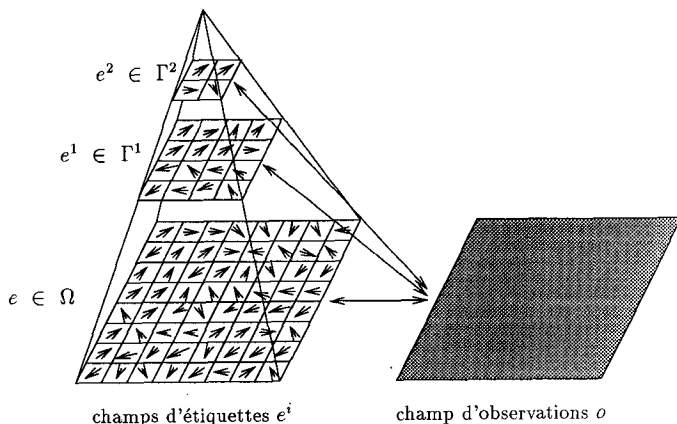


Figure 5. — Structuration pyramidale des primitives.

Il reste à définir pour chaque niveau de résolution i la distribution jointe $p(e^i, o) = \frac{1}{Z^i} \exp \{-U^i(e^i, o)\}$.

Pour cela on poursuit l'identification entre Γ^i et Ω^i . Étant donnée une configuration $e^i \in \Gamma^i$ et $\Phi^i(e^i)$ la configuration correspondante dans Ω^i , on pose :

$$(21) \quad U^i(e^i, o) \triangleq U_1^i(e^i, 0) + U_2^i(e^i)$$

avec

$$(22) \quad U_1^i(e^i, o) \triangleq U_1(\Phi^i(e^i), 0)$$

$$(23) \quad U_2^i(e^i) \triangleq U_2(\Phi^i(e^i)).$$

L'expression (11) permet d'écrire :

$$(24) \quad U_1^i(e^i, o) = \sum_{s \in S^i} W_s^i(e_s^i, o)$$

avec :

$$(25) \quad W_s^i(e_s^i, o) = \sum_{r \in B_s^i} W_r(e_r, o_r)$$

où B_s^i est le bloc de B^i associé au site s de S^i , et e désigne pour simplifier $\Phi^i(e^i)$. On retrouve pour cette énergie l'expression introduite de façon heuristique par Bouman *et al.* dans leur modèle pyramidal [5]. Quant au second terme de l'énergie il s'écrit très simplement d'après (15) :

$$(26) \quad U_2^i(e^i) = \sum_{c \in \mathcal{C}^i} V_c^i(e^i)$$

où les potentiels $V_c^i(e^i)$ se déduisent facilement de (16), (17), (18) et (19). Ils se calculent à partir des potentiels $V_c(e)$.

L'identité d'énergie entre $e^i \in \Gamma^i$ et $\Phi^i(e^i)$ (configuration correspondante dans Ω^i) permet de remplacer de façon équivalente l'estimation des configurations $\hat{e}^i \in \Omega^i$ définies par (8), par l'estimation de la suite de configurations $\hat{e}^i \in \Gamma^i$ définies par :

$$(27) \quad \hat{e}^i \triangleq \arg \min_{e^i \in \Gamma^i} U_1^i(e^i, o) + U_2^i(e^i) \text{ pour } i = 1, \dots, n.$$

Ceci permet une implémentation avantageuse du modèle car dans cette version pyramidale, le volume des données traitées est d'autant plus faible que la résolution est grossière. D'un point de vue pratique, nous traitons ce problème d'optimisation multiple par une classique stratégie descendante associée à un algorithme déterministe (l'algorithme ICM en l'occurrence, [4]) : \hat{e}^n est estimée la première ; ensuite, si \tilde{e}^i désigne l'estimée de \hat{e}^i obtenue, alors l'algorithme d'estimation à la résolution $i-1$ est initialisé par $[\Phi^{i-1}]^{-1} \circ \Phi^i(\tilde{e}^i)$, interpolation par répétition de \tilde{e}^i au niveau $i-1$ (fig. 6).

Notons bien qu'à chaque niveau de résolution, la modélisation markovienne de la distribution jointe $p(e^i, o) = \frac{1}{Z^i} \exp \{-U^i(e^i, o)\}$ se déduit complètement de la distribu-

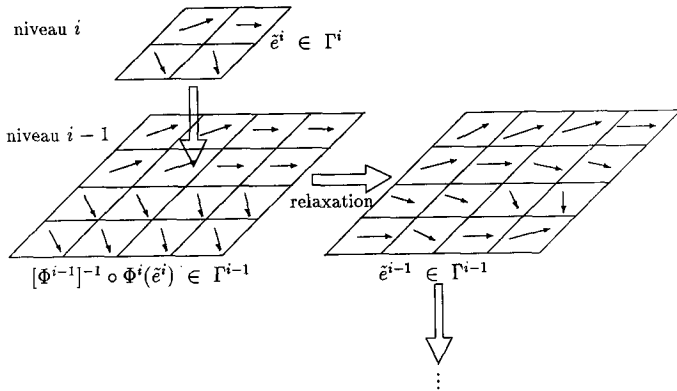


Figure 6. — Stratégie multiéchelle descendante.

tion à la résolution maximale $p(e, o) = \frac{1}{Z} \exp \{-U(e, o)\}$. En particulier on obtient pour la modélisation a priori du champ d'étiquettes E^i un champ de Markov entièrement déterminé (système de voisinage, forme des fonctions de potentiel et paramètres de ces fonctions) par la donnée des caractéristiques du champ à résolution 0.

3. Applications en analyse du mouvement

3.1. APPLICATION À UN MODÈLE DE DÉTECTION DU MOUVEMENT

Nous avons appliqué l'approche multiéchelle qui vient d'être décrite à un modèle markovien de détection du mouvement. Ce modèle permet de détecter le mouvement dans une scène statique avec objets mobiles dans le cas où le recouvrement d'un masque de mouvement par lui-même entre deux images successives est faible [7] (un modèle plus complexe a été développé pour appréhender ce problème de recouvrement [7]).

La grille S considérée est une grille de sites *pixels* et l'on travaille sur une séquence d'images échantillonnées à intervalle dt dans le temps. Soit $f_t(s)$ la luminance (ou intensité) observée à l'instant t au site s (pour plus de commodité le site considéré n'est plus placé en indice pour permettre d'indiquer en indice l'instant considéré). La détection du mouvement repose sur l'analyse des variations de la fonction de luminance entre deux instants $t - dt$ et t . Les observations sont donc les différences inter-images

$$(28) \quad o_t(s) \triangleq |f_t(s) - f_{t-dt}(s)|.$$

A ces observations quantitatives sont ajoutées des observations symboliques binaires $\bar{o}_t(s) \in \{0, 1\}$ indiquant la présence ($\bar{o}_t(s) = 1$) ou l'absence ($\bar{o}_t(s) = 0$) de variations significatives de luminance au site s entre les instants $t - dt$ et t . Elles sont obtenues par un détecteur statistique de changement dans lequel l'intensité est localement modélisée par une fonction linéaire additionnée d'un bruit

blanc gaussien de variance constante (voir [7] pour plus de détails).

Les étiquettes quant à elles prennent leurs valeurs dans $\Lambda = \{0, 1\}$, pour indiquer l'absence ou la présence de mouvement en chacun des sites pixels à l'instant t :

$$(29) \quad \begin{cases} e_t(s) = 0 \Rightarrow s \in \text{zone statique,} \\ e_t(s) = 1 \Rightarrow s \in \text{zone en mouvement.} \end{cases}$$

Enfin les observations aux instants t et $t + dt$ seront nécessaires à la détection des mouvements à l'instant t (on utilise donc les images f_{t-dt}, f_t et f_{t+dt} de la séquence). La fonction d'énergie globale se décompose en trois termes :

$$(30) \quad U(e_t, o_t, o_{t+dt}, \bar{o}_t, \bar{o}_{t+dt}) \triangleq \underbrace{U_{11}(e_t, o_t, o_{t+dt}) + U_{12}(e_t, \bar{o}_t, \bar{o}_{t+dt})}_{U_1(e_t, o_t, o_{t+dt}, \bar{o}_t, \bar{o}_{t+dt})} + U_2(e_t)$$

que nous allons détailler :

- conformément au modèle général présenté en section 2, l'énergie U_2 se décompose en somme de potentiels sur les cliques engendrées par un système de 8-voisinage

$$(31) \quad U_2(e_t) \triangleq \sum_{c \in \mathcal{C}} V_c(e_t).$$

Le choix des potentiels permet de conférer des propriétés a priori sur le champ d'étiquettes. Dans cette application, nous recherchons des masques de mouvement compacts, à l'image des objets mobiles dont ils sont la projection dans le plan image. Nous attribuons pour cela un potentiel non nul aux seules cliques binaires $c = \{s_1, s_2\} \in \mathcal{C}$. Il s'agit d'un potentiel « à niveaux », indépendant de l'orientation des cliques :

$$(32) \quad V_{\{s_1, s_2\}}(e_t) \triangleq \begin{cases} +\beta_2 & \text{si } e_t(s_1) \neq e_t(s_2) \\ -\beta_2 & \text{si } e_t(s_1) = e_t(s_2) \end{cases}$$

où β_2 est un paramètre réel positif.

- U_{11} exprime le lien entre les variations temporelles de luminance dans la séquence (o_t, o_{t+dt}) et le champ d'étiquettes estimé e_t , sachant que la détection des variations temporelles de luminance n'est pas en soi suffisante puisqu'elle produit deux sortes de masques correspondant d'une part aux zones découvertes par l'objet en mouvement et d'autre part celles recouvertes par le même objet (fig. 7 b). L'expression de cette énergie, donnée dans [7] est :

$$(33) \quad U_{11}(e_t, o_t, o_{t+dt}) \triangleq \sum_{s \in S} \left\{ \frac{1}{2\sigma^2} [o_{t+dt}(s) - \Psi(e_t(s))]^2 + \frac{1}{2\sigma^2} [o_t(s) - \Psi(e_t(s))]^2 \right\}$$

où :

$$(34) \quad \Psi(e_t(s)) \triangleq \begin{cases} 0 & \text{si } e_t(s) = 0 \\ \mu & \text{si } e_t(s) = 1 \end{cases}$$

σ^2 est une variance supposée connue et μ est un paramètre réel positif représentant en quelque sorte la variation moyenne de luminance en un site où se produit un mouvement.

• U_{12} permet d'éviter la prise en compte des zones d'occlusion (zones apparaissant ou disparaissant entre deux instants) et de reconstruire ainsi un masque correspondant effectivement à l'objet en mouvement [7] (comparer fig. 7 b et fig. 7 c) :

$$(35) \quad U_{12}(e_t, \bar{o}_t, \bar{o}_{t+dt}) \triangleq \sum_{s \in S} W_\tau(e_t(s), \bar{o}_t(s), \bar{o}_{t+dt}(s))$$

les valeurs de la fonction W_τ étant données par la table 1 (β_1 est un paramètre réel positif).

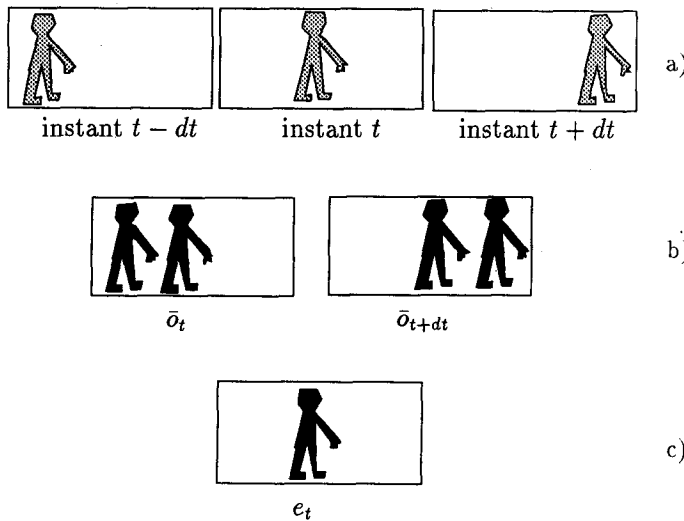


Figure 7. — Détection du mouvement en l'absence de recouvrement des masques.

L'énergie $U_1 = U_{11} + U_{12}$ étant manifestement de la forme (7), ce modèle de détection entre bien dans le cadre général présenté dans la section 2.2. Nous pouvons donc, à partir de ce modèle markovien déduire par la méthode proposée un modèle multiéchelle, dans lequel la distribution jointe à une échelle i est

$$p(e_t^i, o_t, o_{t+dt}, \bar{o}_t, \bar{o}_{t+dt}) = \frac{1}{Z_i} \exp \left\{ -U^i(e_t^i, o_t, o_{t+dt}, \bar{o}_t, \bar{o}_{t+dt}) \right\}$$

où l'énergie globale à l'échelle i s'écrit :

$$(36) \quad U^i(e_t^i, o_t, o_{t+dt}, \bar{o}_t, \bar{o}_{t+dt}) \triangleq U_{11}^i(e_t^i, o_t, o_{t+dt}) + U_{12}^i(e_t^i, \bar{o}_t, \bar{o}_{t+dt}) + U_2^i(e_t^i).$$

D'après (24) et (25) :

$$(37) \quad U_{11}^i(e_t^i, o_t, o_{t+dt}) \triangleq \sum_{s \in S^i} \sum_{r \in B_s^i} \frac{[o_{t+dt}(r) - \Psi(e_t^i(s))]^2 + [o_t(r) - \Psi(e_t^i(s))]^2}{2\sigma^2}$$

$$(38) \quad U_{12}^i(e_t^i, \bar{o}_t, \bar{o}_{t+dt}) \triangleq \sum_{s \in S^i} \sum_{r \in B_s^i} W_\tau(e_t^i(s), \bar{o}_t(r), \bar{o}_{t+dt}(r))$$

la fonction W_τ étant donnée par la table 1.

Table 1. — Définition de la fonction $W_\tau(e_t(s), \bar{o}_t(s), \bar{o}_{t+dt}(s))$

$e_t(s)$	$\bar{o}_t(s)$	$\bar{o}_{t+dt}(s)$	$W_\tau(e_t(s), \bar{o}_t(s), \bar{o}_{t+dt}(s))$
0	0	0	$-\beta_1/2$
0	0	1	$-\beta_1/2$
1	0	0	β_1
1	0	1	β_1
0	1	0	$-\beta_1/2$
0	1	1	β_1
10	10	0	β_1
1	1	1	$-\beta_1/2$

Quant à l'énergie U_2^i , les calculs donnent :

$$(39) \quad U_2^i(e_t^i) \triangleq \sum_{c \in \mathcal{C}^i} V_c^i(e_t^i)$$

où un système de 8-voisinage est considéré. Les potentiels V_c^i sont nuls sur les cliques de plus de deux sites. Si \mathcal{C}_h^i , \mathcal{C}_v^i et \mathcal{C}_d^i sont respectivement les ensembles de cliques binaires horizontales, verticales et diagonales, alors les potentiels sont les suivants :

$$(40) \quad \left\{ \begin{array}{l} \text{pour } s \in S^i \\ \quad V_{\{s\}}^i(e_t^i) \triangleq -p^i \cdot \beta_2 \\ \text{pour } \{s_1, s_2\} \in \mathcal{C}_h^i \cup \mathcal{C}_v^i \\ \quad V_{\{s_1, s_2\}}^i(e_t^i) \triangleq \begin{cases} +q_{hv}^i \cdot \beta_2 & \text{si } e_t^i(s_1) \neq e_t^i(s_2) \\ -q_{hv}^i \cdot \beta_2 & \text{si } e_t^i(s_1) = e_t^i(s_2) \end{cases} \\ \text{pour } \{s_1, s_2\} \in \mathcal{C}_d^i \\ \quad V_{\{s_1, s_2\}}^i(e_t^i) \triangleq \begin{cases} +q_d^i \cdot \beta_2 & \text{si } e_t^i(s_1) \neq e_t^i(s_2) \\ -q_d^i \cdot \beta_2 & \text{si } e_t^i(s_1) = e_t^i(s_2) \end{cases} \end{array} \right.$$

avec $p^i = 2(2^i - 1)(2^{i+1} - 1)$, $q_{hv}^i = 2^i + 2(2^i - 1)$ et $q_d^i = 1$. On observe pour $i > 0$ l'apparition d'une anisotropie spatiale inexistante au niveau le plus fin. En effet, alors que le paramètre associé aux cliques diagonales reste égal à 1, celui correspondant aux cliques horizontales et verticales prend des valeurs qui augmentent avec i .

Le modèle multiéchelle ainsi défini a été appliqué à différentes séquences, associé à une technique descendante classique et à une estimation des champs d'étiquettes par l'algorithme ICM, comme il est expliqué dans la section 2.3. La structuration pyramidale des étiquettes est faite sur trois niveaux ($n = 3$) et l'algorithme ICM est appliqué en balayant l'image dans le sens lexicographique (de gauche à droite et de haut en bas) puis dans le sens contraire. On notera RME (Relaxation Multiéchelle) cette méthode.

À titre de comparaison nous avons appliqué aux mêmes séquences les algorithmes suivants :

- estimation monorésolution du champ d'étiquettes par relaxation déterministe (ICM),
- estimation monorésolution par relaxation stochastique basée sur l'échantillonneur de Gibbs [14]. La loi de descente en température utilisée est exponentielle, i.e. de la forme : $T(j) = T_0 \cdot A^j$ avec $A = 0,97$ et $T_0 = 300$,
- estimation multirésolution classique (stratégie descendante) avec relaxation déterministe (ICM) : on construit une pyramide d'observations par filtrages passe-bas et décimations spatiales d'ordre 2 successifs des images observées. Nous avons utilisé un filtre QMF séparable caractérisé par les coefficients :

$$[0,01995 ; -0,04271 ; -0,05224 ; 0,29271 ; 0,056458 ; 0,29271 ; -0,05224 ; -0,04271 ; 0,01995]$$

L'estimation se fait de la résolution la plus grossière vers la plus fine. L'initialisation au niveau i est faite par simple duplication de l'estimée obtenue au niveau $i + 1$. Le modèle markovien à chaque niveau est celui décrit par (30), (31), (33) et (35), avec les mêmes paramètres pour tous les niveaux,

- estimation multiéchelle sur la base du modèle présenté et défini dans le cadre de la détection par (36), (37), (38), (39) et (40).

Nous noterons RD (Relaxation Déterministe), RS (Relaxation Stochastique), RMR (Relaxation Multirésolution) et RME (Relaxation Multiéchelle) ces quatre algorithmes. Pour les quatre algorithmes, dans le cas où la relaxation n'est pas initialisée par une estimée à un niveau précédent, elle est initialisée par \bar{o}_i^i .

Nous présentons ici l'exemple d'une scène réelle de trafic routier appelée « autoroute » (fig. 8 a) dans laquelle :

- un camion suivi d'une voiture s'éloignent de l'observateur, tandis qu'une seconde voiture est en train de doubler le camion,
- dans le coin supérieur gauche de l'image on aperçoit plusieurs voitures venant en sens inverse.



Figure 8. — Deux images des séquences étudiées : (a) séquence « autoroute », (b) séquence « divergence ».

La taille des images de cette séquence est 128×256 pixels.

Les valeurs données aux différents paramètres sont les mêmes pour les quatre algorithmes et ont été ajustées de façon heuristique (du moins uniquement au niveau 0 en ce qui concerne le modèle multiéchelle) :

$$\beta_1 = 100 \quad \beta_2 = 20 \quad \mu = 10.$$

Le critère d'arrêt est le même dans tous les cas : la relaxation qu'elle soit déterministe ou stochastique est stoppée dès qu'un balayage complet de l'image se produit avec au plus un changement d'étiquette.

Dans ce problème, les énergies obtenues par les différents algorithmes restés sont sensiblement les mêmes (table 2). Cela tendrait à prouver que dans ce cas précis la fonction d'énergie est dépourvue de minima locaux prononcés : quelle que soit la méthode employée, on converge assez rapidement vers le minimum global de l'énergie. Nous indiquons également pour chacun des algorithmes les temps de convergence en temps cpu et en nombre d'itérations équivalentes (table 2). Une itération équivalente correspond au balayage complet de la grille de sites au niveau le plus fin (soit $2^m \times 2^{m'}$ sites). Ainsi pour les algorithmes RMR et RME :

$$nb_{eq} = \sum_{i=0}^n \frac{1}{4^i} nb_i$$

où nb_i est le nombre d'itérations au niveau i . Ceci ne tient donc pas compte du coût calculatoire réel de la remise à jour d'un site, coût qui dépend du niveau de résolution i et de la nature du modèle (multiéchelle ou multirésolution). En particulier la remise à jour d'un site au niveau i ($i > 0$) nécessite plus de calculs dans le modèle multiéchelle en raison du terme énergétique U_i^i liant ce site aux observations des $2^i \times 2^i$ sites correspondant sur la grille initiale. Pour cette raison, nombre d'itérations équivalentes et temps cpu ne sont pas proportionnels.

Table 2. — Temps de convergence (en temps cpu et nombres d'itérations équivalentes) et énergies finales pour la séquence « autoroute »

algorithme	RD	RS	RMR	RME
nb. iter. équivalentes	30	294	17,8	7,25
temps cpu	5 min	49 min	3 min	1 min 29 s
énergie finale	- 4 046 773	- 4 052 312	- 4 046 301	- 4 047 764

Les masques obtenus pour chacun des algorithmes sont présentés figure 9. On constate que la multirésolution classique (RMR) converge plus lentement que l'algorithme multiéchelle et qu'en l'absence d'ajustement des paramètres par rapport à l'échelle, le résultat obtenu est de mauvaise qualité : deux objets mobiles (le camion et la voiture qui le double) n'ont pas été détectés. Cela est dû à l'important recouvrement de ces deux objets entre les deux images successives considérées. Dans la mesure où ils correspondent à des zones assez homogènes, les variations

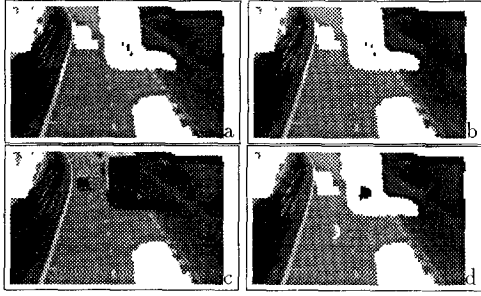


Figure 9. — Résultats des quatre versions de l'algorithme de détection à l'instant t de la séquence « autoroute » : (a) RD (b) RS (c) RMR (d) RME.

temporelles détectées sont très faibles ⁽⁴⁾. Une telle dégradation des résultats par l'introduction de techniques multi-grilles classiques est un risque intrinsèque de cette méthode. En effet, les modèles généralement développés ont un domaine de validité restreint à un type d'images dont les caractéristiques sont rarement indépendantes de l'échelle. Ainsi, dans notre exemple, il est possible d'améliorer les résultats de la relaxation multirésolution (RMR), mais pour cela un réglage entièrement heuristique des paramètres en fonction de la résolution est nécessaire. Cette difficulté n'apparaît pas dans notre relaxation multi-échelle dans la mesure où le modèle intègre lui-même de façon cohérente une adaptation du modèle d'origine aux différents niveaux de la structure pyramidale d'étiquettes. Enfin remarquons que c'est ce dernier algorithme qui présente, dans ce cas, la plus grande vitesse de convergence.

3.2. APPLICATIONS À UN MODÈLE DE MESURE DU DÉPLACEMENT

Nous avons procédé de la même façon pour un modèle classique de mesure du déplacement dans une séquence d'images [20].

Dans ce modèle, les luminances observées, $f_t(s)$, $s \in S$, sont les uniques observations utilisées pour l'estimation. On cherche à estimer le champ de déplacement $\mathbf{d}_t = \{\mathbf{d}_t(s), s \in S\}$ entre les instants t et $t + dt$. Le vecteur $\mathbf{d}_t(s)$ exprime que le point physique situé au site s à l'instant t se trouve au site $s + \mathbf{d}_t(s)$ à l'instant $t + dt$. Ces vecteurs sont donc définis sur la même grille de discrétisation S . On limite de plus l'espace des configurations en imposant aux composantes de ces vecteurs d'appartenir à $\{-d_{\max}, \dots, d_{\max}\}$ avec $d_{\max} \in \mathbb{N}^*$. Ainsi $\Lambda = \{-d_{\max}, \dots, d_{\max}\} \times \{-d_{\max}, \dots, d_{\max}\}$, et $\text{card } \Lambda = (2d_{\max} + 1)^2$.

La fonction d'énergie globale se décompose en deux termes

$$U(\mathbf{d}_t, f_t, t, t + dt) \triangleq \underbrace{\sum_{s \in S} W_s(\mathbf{d}_t(s), f_t(s), f_{t+dt}(s))}_{U_1(\mathbf{d}_t, f_t, f_{t+dt})} + \underbrace{\sum_{c \in \mathcal{C}} V_c(\mathbf{d}_t)}_{U_2(\mathbf{d}_t)}$$

⁽⁴⁾ Un modèle de détection plus sophistiqué prend en compte ce problème de recouvrement dans [7].

où :

- le premier terme correspond à la DFD (différence inter-image déplacée), et exprime l'hypothèse de luminance constante au cours du temps pour un point physique donné :

$$(41) \quad \forall s \in S, \quad W_s(\mathbf{d}_t(s), f_t(s), f_{t+dt}(s)) \triangleq \{f_t(s) - f_{t+dt}(s + \mathbf{d}_t(s))\}^2$$

- le second terme dit de « régularisation » favorise les champs de déplacement « lisses » et pénalise donc les variations brutales entre deux sites voisins. \mathcal{C} étant l'ensemble des cliques engendrées sur S par un système de 8-voisinage :

$$(42) \quad \forall c \in \mathcal{C}, \quad V_c(\mathbf{d}_t) \triangleq \begin{cases} \alpha \|\mathbf{d}_t(s_1) - \mathbf{d}_t(s_2)\|^2 & \text{si } c = \{s_1, s_2\} \\ 0 & \text{sinon} \end{cases}$$

où α est un paramètre réel positif permettant de régler le poids relatif des deux énergies U_1 et U_2 .

Contrairement au modèle de détection précédemment étudié, le présent modèle, bien que de formulation simple, conduit à une optimisation difficile. En effet, du fait de la forte non-linéarité de la fonction à estimer (due au terme de DFD) [21], celle-ci présente de nombreux minima locaux vers lesquels les algorithmes de relaxation déterministe risquent de converger. Il s'agit donc d'un exemple intéressant pour comparer en terme de vitesse de convergence et de qualité de l'estimée, les approches multirésolutions classiques et l'approche multi-échelle que nous proposons.

Le modèle remplissant les différentes conditions énoncées dans la section 2.2, on peut en dériver un modèle multi-échelle caractérisé à l'échelle i par la fonction d'énergie :

$$(43) \quad U^i(\mathbf{d}_t^i, f_t, f_{t+dt}) \triangleq U_1^i(\mathbf{d}_t^i, f_t, f_{t+dt}) + U_2^i(\mathbf{d}_t^i).$$

Comme pour la détection, (24) et (25) donnent :

$$(44) \quad U_1^i(\mathbf{d}_t^i, f_t, f_{t+dt}) \triangleq \sum_{s \in S^i} \sum_{r \in B_s^i} \{f_t(r) - f_{t+dt}(r + \mathbf{d}_t^i(s))\}^2$$

et après calculs, on trouve pour l'énergie U_2^i l'expression suivante :

$$(45) \quad U_2^i(\mathbf{d}_t^i) \triangleq \sum_{c \in \mathcal{C}^i} V_c^i(\mathbf{d}_t^i)$$

où S^i est munie d'une structure de 8-voisinage et les potentiels V_c^i valent :

$$(46) \quad \left\{ \begin{array}{l} \text{pour } c \in \mathcal{C}^i - (\mathcal{C}_h^i \cup \mathcal{C}_v^i \cup \mathcal{C}_d^i) \quad V_c^i(\mathbf{d}_t^i) \triangleq 0 \\ \text{pour } \{s_1, s_2\} \in \mathcal{C}_h^i \cup \mathcal{C}_v^i \\ \quad V_{\{s_1, s_2\}}^i(\mathbf{d}_t^i) \triangleq q_{hv}^i \cdot \alpha \|\mathbf{d}_t^i(s_1) - \mathbf{d}_t^i(s_2)\|^2 \\ \text{pour } \{s_1, s_2\} \in \mathcal{C}_d^i \\ \quad V_{\{s_1, s_2\}}^i(\mathbf{d}_t^i) \triangleq q_d^i \cdot \alpha \|\mathbf{d}_t^i(s_1) - \mathbf{d}_t^i(s_2)\|^2 \end{array} \right.$$

avec $q_{hw}^i = 2^i + 2(2^i - 1)$ et $q_d^i = 1$. Comme dans le cas de la détection du mouvement, l'application du modèle multi-échelle induit pour $i > 0$ une anisotropie spatiale inexistante au niveau le plus fin ($i = 0$).

Pour l'étude comparative des différents algorithmes markoviens d'estimation du déplacement par DFD, nous avons dans un premier temps utilisé deux séquences : la séquence « autoroute » dont il a déjà été question (fig. 8 a) et une séquence dite « divergence » simulant un zoom de caméra sur une image réelle 64×64 de voiture (fig. 8 b). Les quatre algorithmes testés sont, comme pour la détection :

- la relaxation déterministe monorésolution (RD) ;
- la relaxation stochastique (RS) basée sur l'échantillonneur de Gibbs, avec la loi exponentielle de descente en température $T(j) = T_0 \cdot A^j$ ($T_0 = 300$, $A = 0,97$ pour la séquence « divergence » et $T_0 = 500$, $A = 0,98$ pour la séquence « autoroute »).
- la relaxation multirésolution classique (RMR) avec ICM, sur une pyramide à quatre niveaux ($n = 4$), avec la même valeur du paramètre α à tous les niveaux.
- la relaxation multiéchelle (RME) basée sur le modèle précédemment développé, avec $n = 4$.

Pour les deux séquences, nous avons ajusté heuristiquement la valeur de α à 20, et le critère d'arrêt des algorithmes est le suivant : l'algorithme de relaxation est stoppé lorsque un balayage complet de l'image se fait en changeant un nombre d'étiquettes inférieur ou égal à un seuil τ^i dépendant du niveau de résolution ($\tau^0 = 10$, $\tau^1 = 2$, $\tau^2 = 0$ et $\tau^3 = 0$). Enfin, dans le cas où la relaxation n'est pas initialisée par une estimée à un niveau précédent, l'initialisation est aléatoire.

Les temps de convergence (en temps cpu et nombres d'itérations équivalentes) et les énergies finales pour chacun des algorithmes sont fournis dans la table 3 pour la séquence « divergence » et dans la table 4 pour la séquence « autoroute ».

Table 3. — Temps de convergence et énergies finales pour la séquence « divergence ».

algorithme	RD	RS	RMR	RME
nb. iter. équivalentes	37	468	10,78	4,92
temps cpu	43 min 17 s	9 h 7 min 33 s	10 min 12 s	6 min 14 s
énergie finale	139 042	87 292	116 502	89 355

Table 4. — Temps de convergence et énergies finales pour la séquence « autoroute ».

algorithme	RD	RS	RMR	RME
nb. iter. équivalentes	32	442	20,43	9,16
temps cpu	7 h 35 min 15 s	104 h 48 min 7 s	4 h 13 min 57 s	2 h 28 min 50 s
énergie finale	1 117 945	745 999	811 468	749 471

Les courbes de la figure 10 permettent d'apprécier le comportement des différents algorithmes pour la séquence « divergence ». Pour cette séquence, c'est logiquement la relaxation stochastique (RS) qui fournit l'énergie la plus faible, mais au prix d'un nombre d'itérations considérable ; l'algorithme multiéchelle (RME) que nous proposons, tout en étant le plus rapide des quatre, permet d'atteindre une énergie proche du meilleur résultat obtenu par la relaxation stochastique ; enfin, la relaxation multirésolution déterministe classique (RMR) converge rapidement vers un minimum local de valeur assez élevée. Les constats sont strictement les mêmes pour la séquence « autoroute ».

Les champs de déplacement obtenus pour chacune des séquences sont donnés respectivement en figure 11 et figure 12. Ils sont sous-échantillonnés pour plus de lisibilité. L'observation de ces champs va dans le sens des

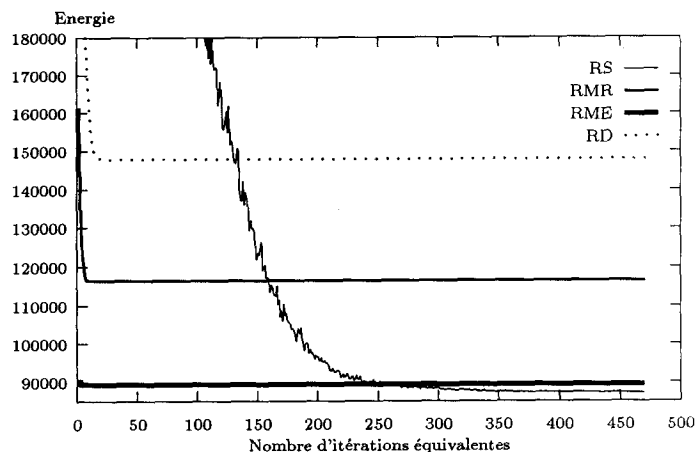


Figure 10. — Énergie U en fonction du nombre d'itérations pour la séquence « divergence ».

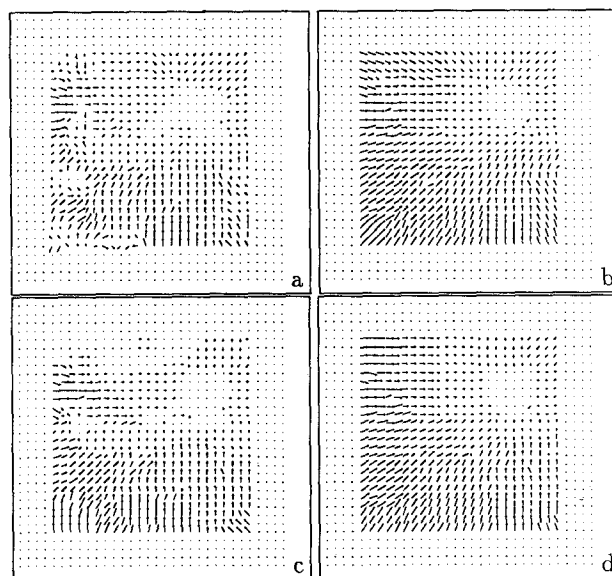


Figure 11. — Champ de déplacement pour la séquence « divergence » : (a) RD (b) RS (c) RMR (d) RME.

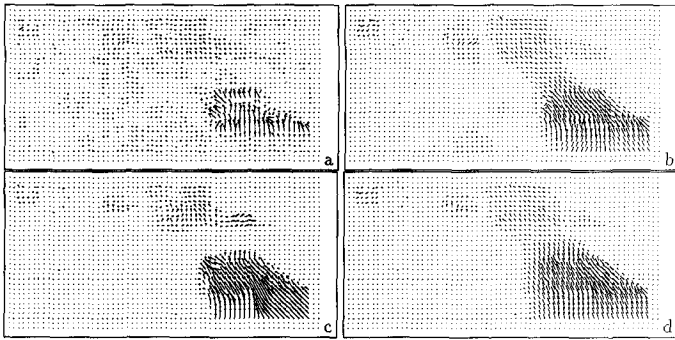


Figure 12. — Champs de déplacement pour la séquence « autoroute » : (a) RD (b) RS (c) RMR (d) RME.

remarques précédentes : l'algorithme multiéchelle proposé fournit des résultats proches de ceux obtenus par une relaxation stochastique très coûteuse en temps de calculs. Pour valider les précédents résultats nous avons testé les quatre algorithmes sur 41 séquences composées de deux images 64×64 pixels obtenues par mouvement synthétique de fragments de scènes réelles. Le mouvement synthétique utilisé est le même pour toutes les paires d'images (fig. 13). Il est composé de quatre régions de mouvement : la région supérieure gauche correspond à une translation, la région supérieure droite est animée d'un mouvement divergeant, la région inférieure gauche est le siège d'une rotation et la région inférieure droite enfin, correspond à la combinaison de la rotation et du mouvement divergent précédents. Enfin la structuration pyramidale des étiquettes est faite sur trois niveaux ($n = 3$).

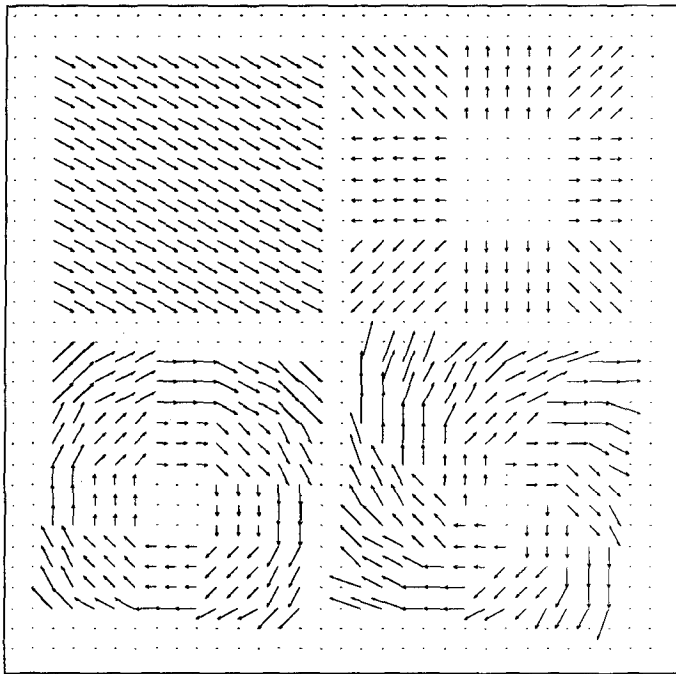


Figure 13. — Champ de déplacement synthétique à estimer sur 41 séquences de deux images.

Le graphique figure 14 fournit pour les 41 paires d'images testées et pour chacun des algorithmes le rapport $\frac{U_{\text{finale}}^{\text{RS}}}{U_{\text{finale}}}$ où U_{finale} est l'énergie de la configuration obtenue par l'algorithme considéré et $U_{\text{finale}}^{\text{RS}}$ est celle obtenue sur la même paire d'images par l'algorithme stochastique. Les graphiques figure 15 et figure 16 donnent quant à eux les temps de convergence des quatre algorithmes en nombres d'itérations équivalentes et temps cpu, l'échelle sur l'axe des ordonnées étant logarithmique. La table 5 donne pour

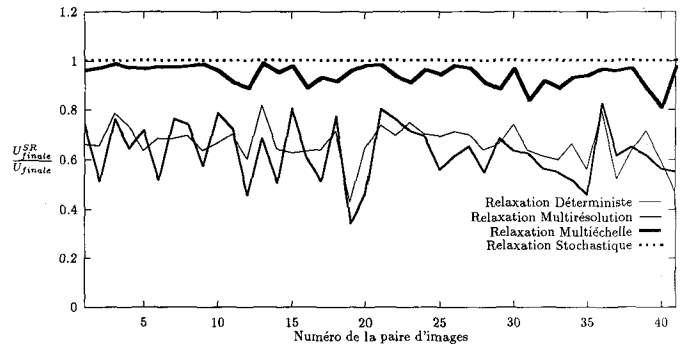


Figure 14. — Rapport $\frac{U_{\text{finale}}^{\text{RS}}}{U_{\text{finale}}}$ pour les 41 séquences de deux images testées.

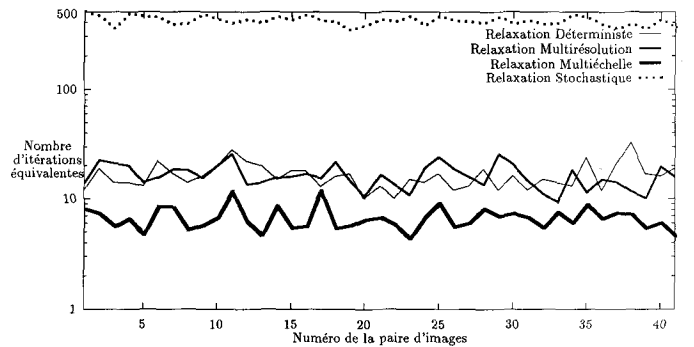


Figure 15. — Temps de convergence en nombres d'itérations équivalentes pour les 41 séquences de deux images testées.

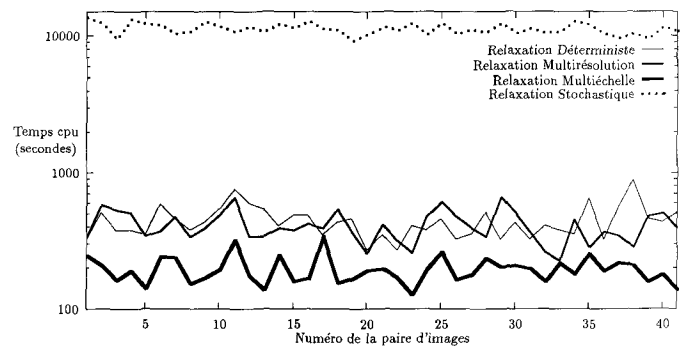


Figure 16. — Temps de convergence en temps cpu pour les 41 séquences de deux images testées.

Table 5. — Temps de convergence moyens et rapports $\frac{U_{finale}^{RS}}{U_{finale}}$ moyens pour le banc test

algorithme	RD	RS	RMR	RME
nb. iter. équivalentes moyen	16,46	409,32	16,32	6,62
temps cpu moyen	7 min 25 s	3 h 4 min 11 s	6 min 48 s	3 min 13 s
$\frac{U_{finale}^{RS}}{U_{finale}}$ moyen	0,663	1,000	0,631	0,940

chacun des algorithmes les valeurs moyennes du temps de convergence et du rapport $\frac{U_{finale}^{RS}}{U_{finale}}$ sur les 41 séquences. Ces

résultats confirment les conclusions tirées des tests précédents : d'un point de vue énergétique, l'algorithme multiéchelle que nous proposons se rapproche du comportement du recuit simulé, moyennant un temps de convergence très inférieur. Visuellement, les champs de déplacement obtenus par relaxation multiéchelle sont proches de ceux produits par relaxation stochastique comme le montre un exemple quelconque tiré de la série de test (fig. 17). La relaxation multirésolution classique quant à elle fournit sur cet ensemble de test des résultats médiocres comparés à ceux de la relaxation multiéchelle.

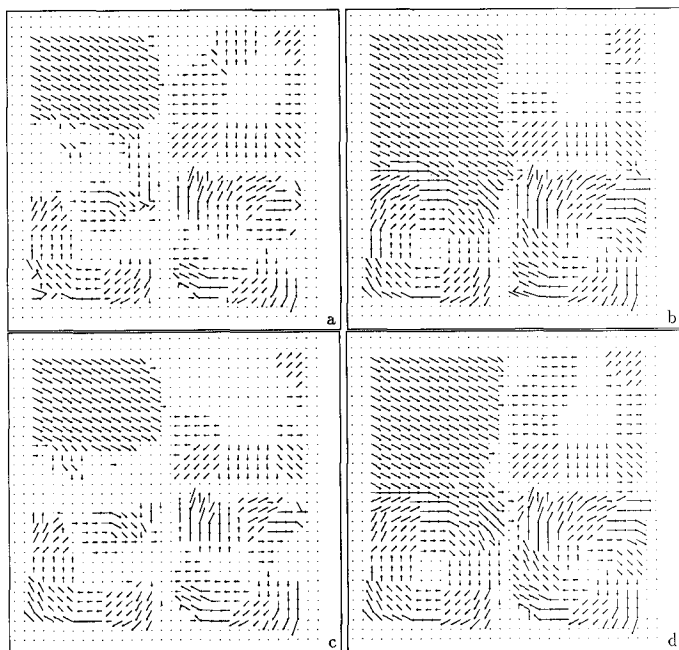


Figure 17. — Exemple de champs des déplacements obtenus sur une des séquences de deux images testées : (a) RD (b) RS (c) RMR (d) RME.

On peut d'ailleurs s'étonner de ce dernier point, à savoir la faiblesse du gain en temps de convergence de la relaxation multirésolution classique sur la relaxation monorésolution. Cela s'explique peut-être par la taille réduite des images initiales (64×64 pixels). Les pyramides d'observations

constituées d'images 32×32 pixels au niveau $i = 1$ et 16×16 pixels au niveau $i = 2$, formées sur les images originales, sont très dégradées eu égard à la taille des filtres utilisés rapportée à la taille des images. En conséquence les champs estimés aux niveaux plus grossiers ($i = 1$ et $i = 2$) sont de qualité moyenne et fournissent au niveau le plus fin ($i = 0$) une initialisation qui n'améliore pas de façon notable la vitesse de convergence ou la qualité du résultat (en terme d'énergie) de la relaxation. N'utilisant que les observations au niveau le plus fin, la relaxation multiéchelle, bien que mettant elle aussi en œuvre une stratégie descendante, s'affranchit de cette difficulté. Notons cependant que l'utilisation d'un filtre passe-bas de support plus court ($[0,25 ; 0,5 ; 0,25]$) ne modifie pas globalement les résultats obtenus par l'algorithme multirésolution classique. Enfin nous pensons que la présence d'un grand nombre de minima locaux pour la fonction à minimiser contribue à dégrader les performances de la méthode multigrille.

Nous avons présenté différents tests sur images réelles animées de mouvements réels ou synthétiques, montrant l'efficacité de l'approche multiéchelle proposée, dans le cas d'une fonction d'énergie présentant de nombreux minima locaux. Tout en mettant en œuvre une relaxation déterministe, elle réussit à éviter ces minima locaux dans lesquels un algorithme multirésolution classique se piège, pour converger très rapidement vers une configuration de bonne qualité. Ceci permet d'envisager avec optimisme l'application de la méthode à des modèles plus complexes pour lesquels la minimisation de la fonction d'énergie est beaucoup plus difficile. En particulier, nous étudions l'implémentation de l'algorithme multiéchelle sur un modèle markovien couplé permettant simultanément l'estimation du champ de vitesse 2D [17] et une segmentation au sens du mouvement [12].

4. Conclusion

Nous avons présenté dans cet article une nouvelle approche markovienne multiéchelle pour le traitement d'images. Nous en avons détaillé le formalisme dans le cadre de la détection et de la mesure du mouvement.

Nous l'avons comparée sur ces deux exemples aux techniques existantes : relaxations déterministe ou stochastique et relaxation multirésolution classique. Nous avons ainsi montré que cette technique multiéchelle présente des vitesses de convergence légèrement supérieures à celle de la technique multirésolution classique à laquelle elle est apparentée par la stratégie de descente dans une structure pyramidale d'étiquettes. Mais contrairement à la méthode classique elle évite les minima locaux de la fonction d'énergie et ne nécessite aucun réglage heuristique de paramètres en fonction du niveau de résolution.

La qualité des résultats fournis par cet algorithme multiéchelle est proche de celle obtenue au prix d'un nombre beaucoup plus élevé d'itérations par un algorithme de relaxation stochastique.

Le modèle markovien multiéchelle que nous proposons présente donc d'intéressantes caractéristiques et de par son degré de généralité il peut être appliqué à d'autres problèmes d'analyse d'images par champs de Markov. Notons enfin que par sa définition particulière des fonctions d'énergies à chaque niveau de résolution, ce modèle peut donner lieu à une parallélisation suivant un schéma d'échange d'information entre niveaux, actuellement à l'étude [24].

Manuscrit reçu le 27 janvier 1992.

Remerciements

Cette étude a été conduite avec le soutien du Groupement de Recherche TdSI (GDR 134) et du PRC « Communication Homme-Machine » (Pôle Vision).

Nous remercions les relecteurs pour leurs commentaires avisés.

BIBLIOGRAPHIE

- [1] J. K. AGGARWAL and N. NANDHAKUMAR, On the computation of motion from sequences of images — a review. — *Proc. IEEE*, Vol. 76, No. 8, p. 917-935, 1988.
- [2] S. T. BARNARD, Stochastic stereo matching over scale. — *Int. J. Comp. Vis.*, Vol. 3, p. 17-32, 1989.
- [3] J. BESAG, Spatial interaction and the statistical analysis of lattice systems. — *J. Royal Statistic. Soc.*, Serie B, Vol. 36, p. 192-236, 1974.
- [4] J. BESAG, On the statistical analysis of dirty pictures. — *J. Royal Statistic. Soc.*, Vol. 48, Serie B, No. 3, p. 259-302, 1986.
- [5] C. BOUMAN and B. LIU, Multiple resolution segmentation of textured images. — *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 13, No. 2, p. 99-113, February 1991.
- [6] P. BOUTHEMY, Extracting dense motion information from an image sequence : Optic flow estimation and related segmentation issues. — In G. R. Taylor, editor, *Kinematic and Dynamic Issues on Sensor Based Control*, volume NATO ASI Series, Vol. F57, p. 223-250. Springer, 1990.
- [7] P. BOUTHEMY and P. LALANDE, Detection and tracking of moving objects based on a statistical regularization method in space and time. — In *Proc. First European Conference on Computer Vision*, p. 307-311, Antibes, France, April 1990. Springer.
- [8] B. CHALMOND, Image restoration using an estimated Markov model. — *Signal Processing*, Vol. 15, No. 2, p. 115-129, September 1988.
- [9] P. B. CHOU and C. M. BROWN, The theory and practice of bayesian image modeling. — *Int. J. Comp. Vis.*, Vol. 4, p. 185-210, 1990.
- [10] P. A. DEVLIVER, Real-time modeling of image sequences based on hidden Markov mesh random field models. — Technical Report M-307, Philips Research Lab., June 1989.
- [11] J. M. DINTEN, Tomographic reconstruction of axially symmetric objects : Regularization by a markovian modelization. — In *Proc. 10th Int. Conf. Pattern Recognition*, Vol. 2, p. 153-158, Atlantic City, June 1990.
- [12] E. FRANCOIS and P. BOUTHEMY, Multiframe-based identification of mobile components of a scene with a moving camera. — In *IEEE Int. Conf. Computer Vision Pattern Recognition*, p. 166-172, Hawaii, June 3-6, 1991.
- [13] D. GEMAN, S. GEMAN, C. GRAFFIGNE and D. PONG, Boundary detection by constrained optimization. — *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 12, No. 7, p. 609-628, July 1990.
- [14] S. GEMAN and D. GEMAN, Stochastic relaxation, Gibbs distributions and the bayesian restoration of images. — *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 6, No. 6, p. 721-741, November 1984.
- [15] B. GIDAS, A renormalization group approach to image processing problems. — *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 11, No. 2, p. 164-180, February 1989.
- [16] W. HACKBUSCH, *Multi-Grid Methods and Applications*. — Springer-Verlag, 1985.
- [17] F. HEITZ and P. BOUTHEMY, Multimodal motion estimation and segmentation using Markov random fields. — In *Proc. 10th Int. Conf. Pattern Recognition*, Vol. 1, p. 378-383, Atlantic City, June 1990.
- [18] F. HEITZ and P. BOUTHEMY, Multimodal estimation of discontinuous optical flow using Markov random fields. — Technical Report 1367, INRIA-Rennes, submitted to IEEE Trans. PAMI, January 1991.
- [19] J. HUTCHINSON, C. KOCH, J. LUO and C. MEAD, Computing motion using analog and binary resistive networks. — *Computer*, Vol. 21, p. 52-63, March 1988.
- [20] J. KONRAD and E. DUBOIS, Multigrid bayesian estimation of image motion fields using stochastic relaxation. — In *Proc. 2nd int. Conf. Computer Vision*, p. 354-362, Tarpon Springs, Florida, December 1988.
- [21] J. KONRAD and E. DUBOIS, A comparison of stochastic and deterministic solution methods in bayesian estimation of 2D motion. — In *Proc. First European Conference on Computer Vision*, p. 149-160, Antibes, France, April 1990. Springer.
- [22] B. C. LEVY, Multiscale models and estimation of discrete Gauss-Markov random fields, communication orale. — In *2nd SIAM Conf. Linear Algebra in Systems Control and Signal Processing*, San Francisco, California, November 1990.
- [23] J. L. MARROQUIN, S. MITTER and T. POGGIO, Probabilistic solution of ill-posed problems in computational vision. — *J. American Statist. Assoc.*, Vol. 82, p. 76-89, 1987.
- [24] E. MEMIN, F. CHAROT and F. HEITZ, Parallel algorithms and architectures for multiscale Markov random field-based image analysis. — In *Workshop on Computer Architecture for Machine Perception*, p. 309-320, Paris, December 1991.
- [25] D. W. MURRAY and H. BUXTON, Scene segmentation from visual motion using global optimization. — *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 9, No. 2, p. 220-228, March 1987.
- [26] P. PEREZ and F. HEITZ, Multiscale Markov random fields and constrained relaxation in low level image analysis. — In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Vol. 3, p. 61-64, San Francisco, March 1992.
- [27] J. SUBRAHMONIA, Y. P. HUNG and D. B. COOPER, Model-based segmentation and estimation of 3D surfaces from two or more intensity images using Markov random fields. — In *Proc. 10th Int. Conf. Pattern Recognition*, Vol. 1, p. 390-397, Atlantic City, June 1990.
- [28] D. TERZOPOULOS, Image analysis using multigrid relaxation methods. — *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 8, No. 2, p. 129-139, March 1986.