Multimodal Estimation of Discontinuous Optical Flow Using Markov Random Fields

Fabrice Heitz and Patrick Bouthemy

Abstract—**The estimation of dense velocity fields from image sequences is basically an ill-posed problem, primarily because the data only partially constrain the solution. It is rendered especially difficult by the presence of motion boundaries and occlusion regions which are not taken into account by standard regularization approaches. In this paper, we present a multimodal approach to the problem of motion estimation in which the computation of visual motion is based on *several complementary constraints*. It is shown that multiple constraints can provide more accurate flow estimation in a wide range of circumstances. The theoretical framework relies on bayesian estimation associated with global statistical models, namely, Markov Random Fields. The constraints introduced here aim to address the following issues: optical flow estimation while preserving motion boundaries, processing of occlusion regions, fusion between gradient and feature-based motion constraint equations. Deterministic relaxation algorithms are used to merge information and to provide a solution to the maximum *a posteriori* estimation of the unknown dense motion field. The algorithm is well suited to a multiresolution implementation which brings an appreciable speed-up as well as a significant improvement of estimation when large displacements are present in the scene. Experiments on synthetic and real world image sequences are reported.**

Index Terms—**Visual motion analysis, discontinuities in optical flow, occlusion processing, multiple constraints, multiresolution analysis, MAP estimate, Markov Random Fields, deterministic relaxation.**

I. Introduction

THE recovery of visual motion from image sequences has motivated number of investigations in the last decade, [2], [26]. The *optical flow field* can be defined as the distribution of 2D velocities of the brightness patterns in the image plane. As optical flow is usually estimated using the spatiotemporal variations of the intensity function within the image sequence, its computed version only imperfectly accounts for the real underlying velocity field due to the relative motion between the camera and the objects in the scene. This problem has been thoroughly addressed by Verri and Poggio, [34], who have shown that the computed *optical flow field* is generally different from the *true 2-D projected motion field* (projection on the image plane of the 3D velocity field of a moving scene). Nevertheless, the discrepancies between the two fields are usually not large, in particular in areas of noticeable intensity gradient values, [34]. Thus optical flow conveys significant information about the 3-D environment, including relative depth, surface orientation, structure and motion of objects in space and sensor motion. Dense optical flow computation thus appears directly relevant to numerous problems in dynamic scene analysis such as moving object detection, motion-based segmentation, [1], [25], [35], qualitative kinematic labeling of moving objects in a scene, [8], recovery of 3D motion and structure, [1], [25] with applications to robot navigation, obstacle avoidance, [8] or image coding.

It is well known that the estimation of dense velocity fields, like many other tasks in low-level vision, is an ill-posed problem. This means that the available data usually do not sufficiently constrain the solution of the problem. Additional smoothness constraints on the resulting motion fields therefore need to be introduced, [18], [27]. Unfortunately, the standard answers to the problem suffer from several shortcomings.

Gradient-based local motion measurements are known to be very sensitive to commonly encountered situations such as regions of constant intensity, motion discontinuities or occlusions areas, [20]. Large displacements are also beyond the scope of those methods. The need for several information sources appears clearly to cope with the variety of real-world images. The usual smoothness constraint also has adverse effects on the estimated optical flow fields because it blurs the motion discontinuities. Among these different problems, the most difficult one is certainly the processing of occlusions between different moving objects in a scene. Occlusions generate discontinuities in the optical flow field and give rise to regions in which no valid motion information is available. The usual computational approaches are not able to cope simultaneously with all these problems, because the constraints they introduce on the desired motion field only imperfectly account for the complexity of real-world scenes.

In this paper, we present a multimodal approach to the problem of motion estimation. The computation of apparent velocity fields is based on several complementary constraints. The constraints introduced here aim at solving the different above-mentioned issues: optical flow estimation while preserving discontinuities, processing of occlusions and introduction of additional information sources. Local motion discontinuities are obtained as a by-product of the method, along with information allowing the algorithm to distinguish occluding regions from occluded ones. New constraints between motion discontinuities, intensity edges and velocity vectors are investigated. For the local motion measurement, two information sources are considered: a gradient-based and a feature-based

Manuscript received November 19, 1990; revised April 21, 1993. This work was supported in part by MRT and CNRS in the context of the PRC Program "Man Machine Interface" under Contract PMFE 88F1 548. Recommended for acceptance by Associate Editor N. Ahuja.

The authors are with IRISA/INRIA, Campus Universitaire de Beaulieu, 35042 Rennes Cedex, France.

IEEE Log Number 9212241.

motion constraint. In our experiments, these two constraints are shown to be complementary. To combine these constraints properly, the validity of each constraint is locally tested using hypothesis tests. A given constraint contributes to the global estimation only if it has been acknowledged as valid.

In our approach, the theoretical and computational framework enabling a cooperation between several sources of information is based on bayesian estimation theory and Markov Random Field (MRF) models. MRF models have been successfully introduced in several important low-level problems of static image processing such as image restoration, [5], [14], image segmentation, [13], stereovision, [3], computed tomography, and surface reconstruction. They have recently been extended to image sequence analysis, for motion detection, [9], motion estimation, [15], [22] and motion-based segmentation, [8], [25]. By defining a coherent mathematical framework for nonlinear global statistical image modeling, they lead to significant improvement with respect to local methods. Markov Random Fields also appear to be an efficient and powerful formalism for specifying spatial interactions between features of a different nature, that is for combining information. MRF modeling allows to jointly handle problems of optical flow estimation and issues of motion discontinuity and occlusion processing. The algorithm combines gradient-based and feature-based velocity measurements with evidence on occlusion areas in order to estimate dense velocity and motion discontinuity maps.

As far as motion estimation is concerned, Markov models were first used by Konrad and Dubois, [21], to estimate discrete-valued velocity vector fields. In [19] Hutchinson *et al.* describe an analog and binary resistive network model equivalent to a Markov Random Field, to perform detection of motion edges simultaneously with the estimation of the velocity field. A VLSI implementation is derived. Konrad *et al.*, [22] and the authors, [15] have independently proposed to introduce binary edge sites between velocity vectors to estimate discontinuous motion fields. A visual motion estimation algorithm, including multiple constraints has also been presented by Black and Anandan in [6]. These constraints include brightness constancy, spatial and temporal coherence. The optical flow field is obtained by minimizing a nonconvex objective function, which can be interpreted as the energy of a MRF model. Motion discontinuities are handled using weak continuity constraints enabling outlier rejection in the optimization scheme.

Here, we propose new comprehensive MRF interaction models for optical flow estimation which differ from the work reported in [6], [15], [19], [22] on the following points.

- The model can integrate different sources of motion measurements. It is illustrated here by a cooperation between gradient-based and edge-based motion measurements, but can be extended to a cooperation with other techniques (correlation, similarity functions, etc.).
- The model properly copes with the problem of discontinuity processing in image sequences. Motion discontinuities are modeled by *local* binary edges located midway between velocity vectors, but an additional feature of the model allows us to take into account *whole regions of discontinuity* and not only the motion boundary lines. These regions correspond to occlusion parts between objects undergoing different motion. This is a key-point since in real world sequences, taking into account false information within an occlusion region may lead to wrong velocity estimates and have an adverse effect on the rest of the velocity field, [2]. It is demonstrated in several examples that the only introduction of *local* binary motion edges is not sufficient to properly handle discontinuities in a moving scene.

The paper is organized as follows. In Section II, we present the two complementary constraint equations: a standard gradient-based constraint [18] and a moving edge constraint, derived from a method presented in [7]. Section III is concerned with the integration of the different constraints within a global bayesian decision framework, based on MRF models. The maximum *a posteriori* (MAP) criterion is adopted. Qualitative as well as quantitative experiments on synthetic and real world images are reported in Section IV. A multiresolution version for our MRF-based relaxation algorithm is described in Section V and Section VI contains concluding remarks.

II. Multiple Motion Constraints

The multimodal motion estimation scheme relies on two motion measurement constraint equations that will be defined hereafter. This differs from the so-called "multiconstraint" methods which rather consider several inputs to the same equation, [24], whereas our multimodal approach is based on a cooperation between different complementary motion constraints. This is a key-point, since problems often arise in the "multiinput" methods because the used inputs (multispectral data for instance) do not supply real complementary information, leading to ill-conditioned systems. The first constraint is the standard motion constraint equation proposed by Horn and Schunck [18], and the second one is related to a moving edge estimation method recently described in [7]. The reliability of these two different motion constraints is discussed and validation factors are associated to both equations, using hypothesis testing techniques.

A. A Gradient-based Motion Constraint

Let $f(x,y,t)$ denote the observed intensity function, where (x,y) designate the 2-D spatial image coordinates and t the time axis. Let $\vec{\omega}_s = (u_s, v_s)^T, (u_s = \frac{dx}{dt}(s), v_s = \frac{dy}{dt}(s))$ denote the velocity vector at point $s = (x,y,t)$. The motion constraint equation is given by, [18]:

$$\vec{\nabla} f(s).\vec{\omega}_s + f_t(s) = 0, \tag{1}$$

where $\vec{\nabla} f$ is the spatial image gradient and f_t stands for $\frac{\partial f}{\partial t}$ and denotes the temporal intensity gradient.

The gradient-based motion constraint relies on the fundamental assumption that the brightness of a moving point is invariant between time t and $t+dt$. This equation shows that only the velocity component parallel to the spatial image gradient can in general be recovered through local computation. This is referred to as the *aperture problem*. In order to derive the

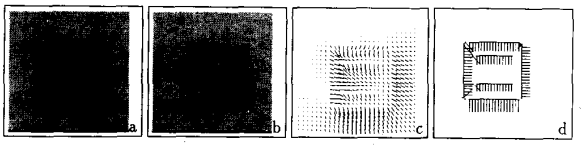


Fig. 1. motion estimation on the *square sequence*. (a)–(b) Original sequence (100 × 100): the square undergoes a translation of (2,2) pixels in the image plane. Gaussian white noise (with variance 4.) has been added to the background. A linearly varying intensity profile has been defined inside the square. (c) Velocity estimation using a standard smoothing method. (d) Estimation of normal velocities using the moving edge estimator described in [7].

complete velocity vector, it is usually assumed that points in the neighborhood of a given point move with similar velocity. *Local optimization approaches* [20], assume constant velocity in the neighborhood whereas *globaloptimizationtechniques* rely on a smoothness assumption of the velocity variations over the whole image, [18].

The limitations of the early gradient-based techniques appear clearly and can be stated as follows. Relation (1) no longer exists in occluded regions, or on motion discontinuities and also on intensity discontinuities (on sharp edges, or in highly textured regions for example). Large displacements are also beyond the scope of these techniques for the same reason. The gradient-based schemes are known to be sensitive to ambiguous areas such as uniform regions or regions exhibiting a linear variation of the intensity in one direction only. Besides, in real world images, the velocity fields are neither locally constant nor globally smooth: they are rather *piece – wise* continuous. In practice, however, the existing schemes show (limited) robustness to these different sources of error, mainly because they minimize some error function with respect to the underlying imperfect model.

The performances of a standard smoothing method based on the image flow constraint equation (1) are illustrated on a synthetic image sequence (Fig. 1(a)–(b)). The sequence exhibits strong intensity discontinuities inside the square, along with occlusion areas corresponding to parts of the background covered or uncovered by the moving pattern. The motion discontinuities lie on the square boundaries. Inside the square the grey value function remains constant along the vertical direction and shows a linear variation along the horizontal one. The standard smoothing method, is close to the one developed by Horn and Schunck [18] and has been derived from our multimodal motion estimator by retaining only the image flow constraint and discarding motion discontinuities and occlusion areas (see Section III). It assumes global smoothness of the flow field. As expected, the resulting velocity field (Fig. 1(c)) is blurred across the motion discontinuities. Moreover, poor quality estimates can be observed both on the central bright intensity lines and in the occlusion areas. These limitations are observed for all standard smoothing methods.

Different solutions have been suggested to cope with some of these problems. The problem of discontinuities in the motion field is considered by Nagel and Enkelmann, [27], who use an *oriented smoothness constraint* which prevents smoothing of the velocity field in directions where significant variations of grey values are detected. Schunck, [29], investigates clustering of local gradient-based constraints in order to obtain homogeneous motion measurements. The detection of motion discontinuities has been considered in several recent papers as a fundamental issue in motion estimation. Two approaches have been studied: the first one detects discontinuities *after* computing the optical flow, [32], the second addresses the detection problem *prior to or simultaneously* with the motion field estimation, [15], [19], [22], [23], [32]. Techniques of the second class give better results because the prior knowledge of motion boundaries helps to prevent velocity smoothing through regions undergoing different movements. Wohn and Waxman, [35] study the global analytic structure of a 2-D motion field and propose a segmentation method based on the recovery of boundaries between regions of analycity in the optical flow field. Experiments on simulated flow fields are presented. Peleg *et al.*, [28] describe a multiresolution approach to extract small moving objects from a static background when camera motion is present. The depth map of the scene is assumed to be known. In [6] motion discontinuities are handled implicitly by using outlier rejection techniques in the estimation of the velocity vector from a small neighborhood. Outlier rejection enables to eliminate measurements which are inconsistent with the local motion, in particular when a motion boundary is present in the neighborhood. Singh [31] has developed an estimation theoretic framework associated with a correlation-based measurement approach which is shown to perform better than conventional smoothing methods at motion boundaries on texture-free images. Besides, Spoerri *et al.*, [32], Little *et al.*, [23] and Black *et al.*, [6] have proposed several occlusion detection techniques based on the analysis of the behavior of matching algorithms in the vicinity of motion boundaries. Multiple motion analysis, of interest in situations involving for instance semi-transparencies, may also be used to detect motion boundaries, when two different motions are observed locally. Contributions in this field are recent, [4], [30] and techniques include separation in the space-time frequency domain, [30] and iterative compensation of multiple movements, [4]. Approaches based on MRF [19], [22] have been mentioned above.

B. An Edge-Based Motion Constraint

We consider here a second complementary motion constraint which is *feature-based*. The underlying estimation method has been described in [7]. It will be called in the subsequent the "moving edge (ME) estimator" and is based on spatio-temporal surface modeling and hypothesis testing techniques. It simultaneously yields from some local processing the following output concerning moving intensity edges: edge position d, edge orientation θ and velocity vector perpendicular to the edge $\vec{\omega_d}^{\perp}$.

To this end, a spatiotemporal edge in an image sequence is modeled as a spatio-temporal surface patch in the (x, y, t) space. Within an elementary volume π in the (x, y, t) space, two local configurations may be encountered: either there is no spatiotemporal edge inside π or there is one. In this case, a surface patch denoted by $S(\Phi)$ subdivides π into two subvolumes π_1 and π_2. Two competing hypotheses H_0 and

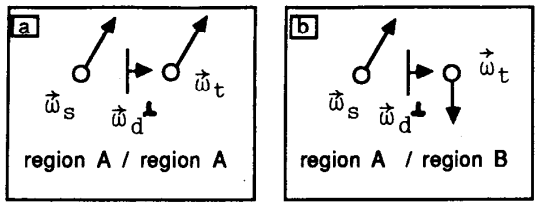


Fig. 2. Location of moving edges with respect to pixels for a vertical edge. (the case of horizontal edges is similar). a) moving edge within a region b) moving edge between two regions

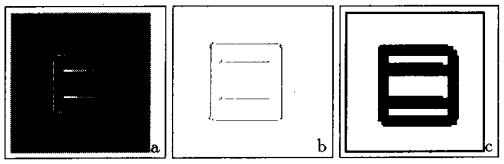


Fig. 3. Confidence factor on the *square sequence* (for the original sequence see Fig. 1). (a) Moving edge constraint: log-likelihood surface for the ME estimator. (b) Moving edge constraint: Binary confidence factor $\xi_{me}(d)$ (black points correspond to $\xi_{me}(d) = 1$). (c) Image flow constraint: Confidence factor ξ_g (black points correspond to $\xi_g = 0$).

H_1 are associated to these configurations, [7] and formalized by the corresponding likelihood functions. Intensities within π are assumed to be independent Gaussian random variables whose mean depends on the considered hypothesis.

A log-likelihood ratio test is designed for a predefined set of values of the parameters Φ describing surface S. Planar patches are considered. Computation mainly reduces to local convolution operations. Edge location, orientation and displacement are directly related to the optimal parameters $\hat{\Phi}$ of the determined planar surface patch if present. To decide whether a ME is present or not, the log-likelihood ratio is compared to a threshold λ (we refer the reader to [7] for more details). The local motion measurement is reliable, even on motion discontinuities and for important displacement magnitudes. Due to the aperture effect only the perpendicular component of the displacement can be derived.

In the ME estimator version we use, edge sites d can be considered as located midway between pixel sites (Fig. 2). To save computation time a spatial intensity edge detection, [10] is performed to determine edge locations. The ME estimator is only applied at these locations to estimate vector $\vec{\omega_d}^{\perp}$.

The velocity component perpendicular to the edge at location d constrains the unknown nearest neighbor vectors $\vec{\omega_s}$ and/or $\vec{\omega_t}$ through a projection equation called *the moving edge constraint equation* (the notations refer to Fig. 2):

$$\vec{\omega_z}.\frac{\vec{\omega_d}^{\perp}}{\| \vec{\omega_d}^{\perp} \|} - \| \vec{\omega_d}^{\perp} \| = 0, \qquad (2)$$

where $\vec{\omega_z}$ designates vectors $\vec{\omega_s}$ or $\vec{\omega_t}$, and $\frac{\vec{\omega_d}^{\perp}}{\|\vec{\omega_d}^{\perp}\|}$ is the unit vector normal to the intensity edge. The moving edge constraint states that the projection of the unknown velocity $\vec{\omega_z}$ on the unit vector perpendicular to the moving edge is equal to the norm of the perpendicular component $\vec{\omega_d}^{\perp}$.

If the detected moving intensity edge is related to an occlusion between two different regions, the constraint only holds for the velocity vector belonging to the same region as the occluding edge. In the case of Fig. 2(b) for instance, it turns out that the constraint only holds for vector $\vec{\omega_s}$, and not for vector $\vec{\omega_t}$. In order to propagate the constraint properly, it is necessary to determine to which region the occluding edge belongs. Information allowing the algorithm to differentiate between the occluding region and the occluded one is therefore required. This information will be defined later in the global markovian modeling (Section III-A). It will be considered as an additional, unknown feature in the model, and will be estimated in parallel with velocities and motion discontinuities.

Fig. 1(d) presents the result of the moving edge estimator on the *square sequence*. One can point out that unlike the gradient-based approach (Fig. 1(c)) the moving edge estimator yields good measurements on motion and intensity discontinuities (however motion information remains obviously sparse and only perpendicular velocity components are recovered).

C. Reliability of Motion Constraints

The accuracy and the reliability of the partial measurements associated to motion constraints (1) and (2) depend on the adequacy between the observed variations of the intensity pattern and the spatiotemporal changes the underlying model accounts for. For the different existing motion constraint equations, little attention has been given to getting reliability measurements (apart from [7], [20]). In the following we define validation factors associated to the constraints we have introduced. They will be used to withdraw the contribution of invalid local constraints from the global estimation.

A Validation Factor for the Moving Edge Constraint: As far as the moving edge constraint is concerned, the ME estimator provides a natural way to define a validation factor. Let us recall that the determination of a moving edge at location d leads to comparison of the log-likelihood ratio associated with the two competing hypotheses to a threshold (see Section II-B or [7]). The optimal value of the log-likelihood ratio L_d at location d, with respect to surface parameters $\hat{\Phi}$ can be used to measure the reliability of the corresponding moving edge.

Fig. 3(a) shows the log-likelihood surface in the case of the moving square. The likelihood surface indicates high reliability in the vicinity of edge location. We introduce following binary-valued validation factor ξ_{me}:

$$\begin{aligned} \xi_{me}(d) &= 1, \qquad \text{if} L_d(\hat{\Phi}, \hat{c}_0, \hat{c}_1, \hat{c}_2) > \lambda_1, \\ \xi_{me}(d) &= 0, \qquad \text{else}, \end{aligned} \qquad (3)$$

where λ_1 is a threshold. The validation factor then gives the location of the most reliable moving edges (Fig. 3(b)).

A Validation Factor for the Image Flow Constraint: Basically the image flow constraint (1) holds as long as the local intensity variations and the observed temporal changes are related. This property refers to the local spatiotemporal linearity and derivability of the intensity function.

To determine whether the observed variations preserve the spatiotemporal relation, it is shown in [16] that it is sufficient in practice to test if the first order spatial derivatives of the intensity function at point s remain the same between time t and $t+1$. The hypothesis test considered here relies on a local

linear model for intensity function $f(x, y, t)$ at point $s = (x, y)$ in two successive images (t and $t + 1$) of the sequence:

$$f(x + \delta x, y + \delta y, t) = f(x, y, t) + a_t \delta x + b_t \delta y + n_1 \quad (4)$$

$$f(x + \delta x, y + \delta y, t + 1) = f(x, y, t + 1) + a_{t+1} \delta x + b_{t+1} \delta y + n_2, \quad (5)$$

where n_1 and n_2 are assumed to be independent zero-mean Gaussian noises with the same variance σ^2.

The reliability of the constraint equation is tested by considering two competing hypotheses denoted H_0 and $\bar{H}_0$, where

$$H_0 : \{a_t = a_{t+1} \text{ and } b_t = b_{t+1}\} \text{ in } W(s)$$
$$\bar{H}_0 : \{a_t \neq a_{t+1} \text{ or} b_t \neq b_{t+1}\} \text{ in } W(s). \quad (6)$$

$W(s)$ designates a local window centered at point s, in which the parameters are estimated. The likelihood functions under each hypothesis are computed, assuming gaussian noises with same variances for n_1 and n_2 and the log-likelihood ratio is compared to a threshold. As in the case of the ME constraint, we define a binary validation factor for the motion constraint at site s:

$$\xi_g(s) = 1, \quad \text{if } H_0 \text{ is selected} \quad (7)$$
$$\xi_g(s) = 0, \quad \text{if } H_0 \text{ is rejected.}$$

Fig. 3(c) shows the sites corresponding to $\xi_g(s) = 0$ issued from test (7) on the *square sequence*. Let us notice that these regions are closely related to the occlusion areas (near the square boundaries) and to spatial discontinuities in the intensity pattern (along the two central lines). In these different areas the image flow constraint is indeed invalid: in the following, ξ_g will help us to disregard these wrong local constraints.

As intuitively expected comparing Fig. 3(b) and Fig. 3(c), one can see that the two motion constraints are complementary in this example. Experiments with other sequences have given similar results: the gradient-based constraint is invalid in regions of "spatiotemporal discontinuity" (within occlusions, on sharp intensity edges and in highly textured regions) whereas the ME method is reliable at those locations (excepted in textured regions). This experimental statement justifies the use of these two particular motion constraints.

III. A Global Bayesian Formulation For Multimodal Motion Estimation

Global bayesian estimation defines a coherent mathematical framework to extract labels describing motion from image sequences. The estimation process can be outlined as follows.

- One or more specialized low-level modules extract from the image sequence features (gradients, moving edges, etc.) that will be used as observations in the estimation process.
- Observations are combined within local photometric and structural models with a priori generic knowledge on the expected result, in order to derive estimates of the unknown labels.

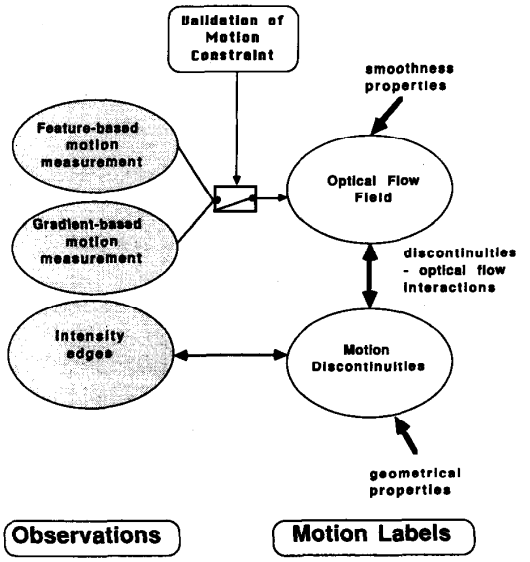


Fig. 4. The multimodal interaction model: interactions between motion labels and observations. Thanks to the MRF model, the motion discontinuities are estimated jointly with the optical flow field. The motion measurement relies on two complementary motion constraints: gradient and feature-based. The reliability of local motion constraints is tested and they contribute to the global estimation only if they are valid. The estimation of motion discontinuities is supported by intensity edges. Geometrical constraints are also defined on the desired discontinuity configurations.

The spatial interactions between observation fields and motion labels are specified using Markov random field (MRF) models. The random field models are employed to provide constraints on the solution and to fuse information.

In the MRF model designed here, local motion discontinuities are simultaneously estimated with the velocity field and multiple local constraints contribute to the estimation of those fields. Intensity edges are used as additional evidence to support the estimation of motion boundaries (Fig. 4).

A. Observations and Labels Supporting Motion Information

In the estimation process, information about motion is summarized in the following labels.

- Vectorial labels $\vec{\omega}_s, (\vec{\omega}_s \in \mathbb{R}^2)$ corresponding to the velocity field $\vec{\omega} = \{\vec{\omega}_s, s \in S\}$ where S denotes the set of pixel sites in the image plane. A local velocity vector is thus associated to every point s in the image plane.
- A set of discrete labels $\gamma = \{\gamma_d, d \in D\}$ describing local motion discontinuities. D denotes the set of edge sites located midway between the pixel sites. There are three possible states for motion discontinuities: $\gamma_d = 0, 1$ or -1.

  $\gamma_d = \pm 1$ (resp. $\gamma_d = 0$) means that a motion discontinuity (resp. no motion discontinuity) is present at location d. In case that $\gamma_d = \pm 1$ the sign of γ_d codes the relative position of the occluding surface with respect to the motion discontinuity, (see Fig. 5). $\gamma_d = +1$ (resp. $\gamma_d = -1$) means that for a vertical discontinuity the occluding region is on the right-hand side (resp. on the

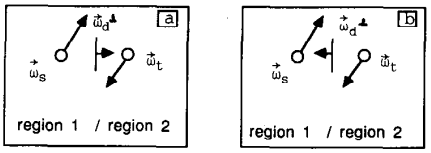


Fig. 5. The sign of label γ_d defines the region to which the occluding edge belongs (case of a vertical edge) (a) Case $\gamma_d = -1$: the occluding edge belongs to region 1 ($\vec{\omega_d}^{\perp}$ and $\vec{\omega}_s$ are consistent). (b) Case $\gamma_d = +1$: the occluding edge belongs to region 2 ($\vec{\omega_d}^{\perp}$ and $\vec{\omega}_t$ are consistent).

left-hand side). A similar code is used for horizontal discontinuities. This three state edge description is used to propagate the moving edge constraint to the proper region (see Section II-B), but can also be considered as a useful by-product of the estimation scheme.

Observations correspond to the output of four independent modules.

- A first module computing the spatial and temporal derivatives $\mathcal{D}f = \{\vec{\nabla} f(s), \frac{\partial f}{\partial t}(s), s \in S\}$ of the intensity function f at every vector site s.
- The moving edge estimation module described in Section II-B, which yields displacement information about intensity edges located on grid D (only the displacements perpendicular to the edges are determined). The set of sites $d \in D$, for which a moving edge exists, will be denoted D_{me}. The corresponding local motion measurement set is denoted $\omega^{\perp} = \{\vec{\omega_d}^{\perp}, d \in D_{me}\}$.
- The validation factors defined in Section II-C, $\xi_g = \{\xi_g(s), s \in S\}$, $\xi_g(s) \in \{0,1\}$ and $\xi_{me} = \{\xi_{me}(d), d \in D\}$, $\xi_{me}(d) \in \{0,1\}$. Observations ξ_g state, at every pixel location s, whether the image flow constraint is reliable or not (the same holds for ξ_{me} at edge locations d for the moving edge constraint).
- A spatial intensity edge detector derived from Canny's criterion proposed by Deriche, [10], provides a binary information about intensity discontinuities on sites $d \in D$. $\eta = \{\eta_d, d \in D\}$ designates the binary map output of the intensity edge detector ($\eta_d = 1$ if an intensity edge is detected). Following [12] intensity edges will be used as partial evidence supporting the state of motion discontinuities γ_d at the same locations.

As explained in Section II-B, the output of the intensity edge detector is also used in the moving edge estimator to reduce the global computation cost. The spatial locations of the moving edges $\omega^{\perp}$ thus coincide with the locations of the intensity edges used to support motion boundary detection.

B. Global Bayesian Decision and the MAP Criterion

The maximum *a posteriori* (MAP) criterion has been widely used in the context of global bayesian decision, [9], [14], [22], [25]. To derive the unknown label fields $(\vec{\omega}, \gamma)$ from the observed fields $(\mathcal{D}f, \omega^{\perp}, \xi_g, \xi_{me}, \eta)$, the following optimization problem has to be solved:

$$\max_{\{\vec{\omega},\gamma\}} p(\mathcal{D}f, \omega^{\perp}, \xi_g, \xi_{me}, \eta, \vec{\omega}, \gamma), \tag{8}$$

where $p(\mathcal{D}f, \omega^{\perp}, \xi_g, \xi_{me}, \eta, \vec{\omega}, \gamma)$ is the joint distribution of the observed and hidden variables.

The distribution of observations and motion labels are specified using a coupled Markov Random Field (MRF) model whose distribution is written in the following form ([14]):

$$p(\mathcal{D}f, \omega^{\perp}, \xi_g, \xi_{me}, \eta, \vec{\omega}, \gamma) = \frac{1}{Z} exp - U(\mathcal{D}f, \omega^{\perp}, \xi_g, \xi_{me}, \eta, \vec{\omega}, \gamma)$$

where

$$U(\mathcal{D}f, \omega^{\perp}, \xi_g, \xi_{me}, \eta, \vec{\omega}, \gamma) = \sum_{c \in C} V_c(\mathcal{D}f, \omega^{\perp}, \xi_g, \xi_{me}, \eta, \vec{\omega}, \gamma) \tag{9}$$

is called the energy function of the MRF. The lowest energies correspond to the most likely configurations. Such a formulation is possible since we assume that the interactions between the different variables remain local, with respect to a chosen neighborhood system ν (see [14] for a complete theory of MRF). C denotes the set of cliques associated to neighborhood system ν. Cliques c are subsets of sites which are mutual neighbors. The potential function V_c is locally defined on clique c and expresses the local interactions between the different variables of the clique. The form of the potential functions is of course problem dependent. The functions that we have defined for the motion estimation scheme integrate the different modeling aspects already discussed: regularization of the velocity field along with preservation of motion boundaries, multimodal cooperation between different measurement sources, discarding of invalid local motion constraints (in particular in the occlusion regions), processing of motion discontinuities... Within this framework, finding the maximum *a posteriori* estimate amounts to minimization of the global energy function U.

The neighborhood system ν is defined on sets S and D as explained in Fig. 6(a). Interactions between observations, velocities, and motion discontinuities are supported by mixed cliques, whereas edge cliques support the geometric properties of motion discontinuities (Fig. 6(b)).

Interactions between variables are modeled through the following decomposition of the global energy function:

$$\begin{aligned} U(\mathcal{D}f, \omega^{\perp}, \xi_g, \xi_{me}, \eta, \vec{\omega}, \gamma) &= U_1(\mathcal{D}f, \xi_g, \vec{\omega}) + U_2(\omega^{\perp}, \xi_{me}, \vec{\omega}, \gamma) \\ &\quad + U_3(\vec{\omega}, \gamma) + U_4(\eta, \gamma) + U_5(\gamma). \end{aligned}$$

Each term is decomposed into local potential functions defined on the different cliques (for the notations see Fig. 5 and Fig. 6) in the equations at the bottom of the next page where Φ_{α_3} is given by

$$\Phi_{\alpha_3}(\| \vec{\omega_s} - \vec{\omega_t} \|) = \text{sign}(\| \vec{\omega}_s - \vec{\omega}_t \| - \alpha_3) \frac{(\| \vec{\omega}_s - \vec{\omega}_t \| - \alpha_3)^2}{{\alpha_3}^2},$$

the $\alpha_i, i = 1, \cdots, 4$, are model parameters and the $C_i, i = 1, \cdots, 8$ denote the different clique types depicted in Fig. 6(b).

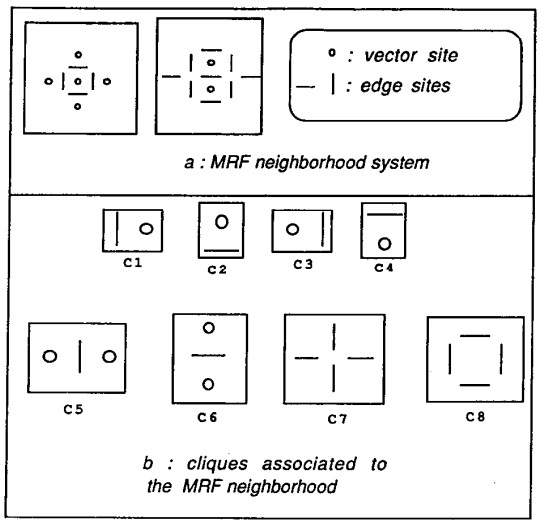


Fig. 6. MRF neighborhood system including vector and edge sites.

In this decomposition, each term expresses a different interaction model, each of which contributes to the global estimation process. The different terms can be interpreted as follows.

Energies U_1 and U_2 (Motion Measurement Constraints): Energies U_1 and U_2 are related to the motion constraint equations (1) and (2) upon which the motion estimation is based. Those energy functions take into account the confidence factors discussed in Section II-C. For sites s belonging to smooth spatiotemporal regions ($\xi_g(s) = 1$), the image flow constraint equation is applied (energy U_1), whereas in sites where $\xi_{me}(s) = 1$, (i.e. presence of a moving edge), the moving edge constraint is considered (energy U_2).

If the moving edge at site d corresponds to a motion discontinuity $|\gamma_d| = 1$, the moving edge constraint is only propagated to the proper region depending on the sign of γ_d. For instance for a vertical motion boundary (see Fig. 5), when $\gamma_d = 1$, $\frac{1}{2}(-|\gamma_d| - \gamma_d + 2) = 0$ and $\frac{1}{2}(-|\gamma_d| + \gamma_d + 2) = 1$; hence, the moving edge constraint is only propagated to the vector $\vec{\omega}_t$ belonging to the right-hand side region (Fig. 5). When $\gamma_d = 0$, i.e., when there is no motion discontinuity at site d, the moving edge constraint is propagated to the regions on both sides.

Conversely, when the velocities are given, two configurations may be encountered.

- The measured orthogonal component $\vec{\omega_d}^{\perp}$ is exactly consistent with the neighboring velocities $\vec{\omega_s}$ and $\vec{\omega_t}$, that is $\vec{\omega_t}.\frac{\vec{\omega_d}^{\perp}}{\|\vec{\omega_d}^{\perp}\|} - \vec{\omega_d}^{\perp} = \vec{\omega_s}.\frac{\vec{\omega_d}^{\perp}}{\|\vec{\omega_d}^{\perp}\|} - \vec{\omega_d}^{\perp} = 0$. In this case, energy U_2 has no effect on the choice of γ_d.
- A discrepancy exists between $\vec{\omega_d}^{\perp}$ and its neighbor vectors. In this case, energy U_2 favors either value $\gamma_d = -1$ or $\gamma_d = +1$ according to the lowest local energy. For a vertical discontinuity for instance, if the perpendicular component $\vec{\omega_d}^{\perp}$ is consistent with vector $\vec{\omega}_s$ (Fig. 5), the value -1 is assigned to γ_d. This means that the occluding surface is on the left-hand side of the boundary in this case. The occluding region is the one containing the velocity vector ($\vec{\omega}_s$ or $\vec{\omega}_t$) the most consistent with the observed component $\vec{\omega_d}^{\perp}$.

Let us point out that energy function U_2 does not decide between states $\gamma_d = 0$ and $|\gamma_d| = 1$, i.e. whether a motion boundary has to be introduced or not. Energy U_2 only makes a selection among states $\gamma_d = 1$ and $\gamma_d = -1$, that is, assigns an existing motion boundary to the region which it belongs to. The decision for the placement of a motion boundary at a given location is inferred from energy term U_3, based on the variations in the estimated velocity field.

Energy U_3 (Velocity Field Smoothing While Preserving Discontinuities): Interactions between velocities and motion discontinuities are supported by mixed cliques (s,t,d). The chosen potential function smooths out the motion field using terms of the form $\| \vec{\omega}_s - \vec{\omega}_t \|$, corresponding to first order derivatives (second order terms were also used, but appeared more sensitive to noise). Velocity smoothing is inhibited when a motion boundary is present ($|\gamma_d| = 1$). Conversely when

$$
\begin{aligned}
U_1(\mathcal{D}f, \xi_g, \vec{\omega}) &= \alpha_1 \sum_{s \in S} \xi_g(s)(\vec{\nabla} f(s).\vec{\omega}_s + f_t(s))^2 \\
U_2(\omega^{\perp}, \xi_{me}, \vec{\omega}, \gamma) &= \\
&\alpha_2 \sum_{s \in S, d \in D_{me}, (s,d) \in \{C_3\} \cup \{C_4\}} \xi_{me}(d)(\vec{\omega_s}.\frac{\vec{\omega_d}^{\perp}}{\| \vec{\omega_d}^{\perp} \|} - \| \vec{\omega_d}^{\perp} \|)^2 \frac{1}{2}(-|\gamma_d| - \gamma_d + 2) \\
&+ \alpha_2 \sum_{t \in S, d \in D_{me}, (t,d) \in \{C_1\} \cup \{C_2\}} \xi_{me}(d)(\vec{\omega_t}.\frac{\vec{\omega_d}^{\perp}}{\| \vec{\omega_d}^{\perp} \|} - \| \vec{\omega_d}^{\perp} \|)^2 \frac{1}{2}(-|\gamma_d| + \gamma_d + 2) \\
U_3(\vec{\omega}, \gamma) &= \sum_{(s,t,d) \in \{C_5\} \cup \{C_6\}} \Phi_{\alpha_3}(\| \vec{\omega_s} - \vec{\omega_t} \|)(1 - |\gamma_d|) \\
U_4(\eta, \gamma) &= \alpha_4 \sum_{d \in D} (1 - \eta_d)|\gamma_d| \\
U_5(\gamma) &= \sum_{c \in \{C_7\} \cup \{C_8\}} V_c(\gamma),
\end{aligned}
\tag{10}
$$

the vector field shows important variations in edge vicinity ($\| \vec{\omega_s} - \vec{\omega_t} \| > \alpha_3$), the placement of a motion discontinuity at that site is favored ($|\gamma_d| = 1$).

Energy U_4 (Interactions Between Motion Discontinuities and Intensity Edges): Intensity edges are used as partial evidence for the determination of motion boundaries. In real world scenes, a 3-D configuration resulting in a motion discontinuity generally also contributes to an intensity edge. Hence, following [12], we assume that motion discontinuities appear with a rather low probability when there is no intensity edge at the same location. This is specified using energy function U_4 with a large positive value for parameter α_4. U_4 prevents configurations in which $|\gamma_d| = 1$ and $\eta_d = 0$ from appearing. This can be a limitation in situations in which the motion boundaries are not supported by underlying intensity edges (two random dot patterns, one occluding the other, for instance). Such (rather uncommon) situations are beyond the scope of the method.

Energy U_5 (Edge Geometry): The edge cliques considered in U_5 help to discourage undesirable geometric configurations (edge ending, isolated or double edges [14]). Two main methods for specifying the geometrical properties of edges in MRF have been proposed, [13,14]. The first approach, [14], consists of assigning different weights to the different possible local edge configurations defined on the geometrical cliques. A large weighting on a configuration tends to discourage this configuration, [14]. The main drawback of this method is to introduce an important number of additional parameters in the model, corresponding to the different weights. The tuning or learning of those parameters is generally not an easy task. As we are concerned with a three state discontinuity description and with the chosen neighborhoods, the number of local configurations is as large as 162. Therefore we have resorted to an alternate approach recently described in [13]. This other solution consists of introducing *forbidden* edge configurations. Each forbidden local configuration (edge endings, isolated or double edges, impossible configurations of occluding/occluded regions), weighs an elementary weight of 1 in energy U_5. The following constrained optimization problem is then solved:

$$\min_{\{\vec{\omega},\gamma\}} (U_1 + U_2 + U_3 + U_4) + \alpha_5 U_5 \text{ with } \alpha_5 = +\infty. \quad (11)$$

A constrained optimization may be obtained in practice by letting $\alpha_5 \nearrow +\infty$ (see [13] for convergence theorems).

C. Energy Minimization Using Deterministic Relaxation

Finding the MAP estimate of the fields $\vec{\omega}$ and γ is equivalent to minimizing the global energy function $U(\mathcal{D}f, \omega^{\perp}, \xi_g, \xi_{me}, \eta, \vec{\omega}, \gamma)$. This global energy function depends both upon continuous and discrete valued variables ($\vec{\omega}$ and γ). To reach configurations close to the global minimum of an energy function, stochastic optimization methods for continuous and discrete variables have been studied [14]. Stochastic optimization algorithms are very time consuming, especially for continuous variables. Most of the recent papers resort to deterministic schemes which are more appealing, as far as computation time is concerned. Deterministic relaxation converges to a local minimum of the energy function, but this loss of optimality may be compensated for by an appropriate initial guess. Besides, in many cases the suboptimal solution can be considered as a relevant solution. In our experiments, we use a modified version of Iterated Conditional Modes [5], a deterministic alternative to simulated annealing. In this relaxation scheme the final result depends on initialization and site visiting order. Satisfactory results are obtained by initializing vectors with $\vec{0}$ and motion boundaries with the intensity edges map η_d. This suggests that for the optimization problem at hand, the initial guess for the velocity field is not so critical.

In the Iterated Conditional Modes relaxation method, the global energy function (10) is minimized by sequentially updating the different sites of the velocity and discontinuity fields. The site visiting order is raster scan, with reverse order after every full sweep of the image. Vector sites and discontinuity sites are visited in turn. At a given location, the label value assigned to a site is the one maximizing the decrease of the global energy function. Thanks to the decomposition of the MRF into local interaction terms, it turns out that updating a site leads to the minimization of a local energy function that depends on the visited site and its neighbors, [14].

The terms of the local energies are derived from the global one (10), apart from one exception concerning the smoothing term $\Phi_{\alpha_3}(\| \vec{\omega_s} - \vec{\omega_t} \|)$, which has been replaced by the simplified quadratic form $\alpha_3' \| \vec{\omega_s} - \vec{\omega_t} \|^2$ in the local energy function $E(\vec{\omega}_s)$ used to update $\vec{\omega_s}$ (the expression of the local energy functions as well as other implementation details may be found in [16]). Using such a simplified form makes the computation of the minimum of $E(\vec{\omega}_s)$ easier since it becomes quadratic with respect to $\vec{\omega}_s$. No visible degradation on the final velocity fields was noticed when this simplification was used.

The other model parameters are either set to a fixed value, whatever the sequence at hand, or computed from the other parameters. We have taken $\alpha_1 = 1, \alpha_2 = 1, \alpha_4 = \frac{4}{{\alpha_3'}^2}$. Indeed, the only parameters that need to be tuned are α_3' and α_3 which weight the interactions between smoothing of velocities and introduction of motion discontinuities. α_3' and α_3 have been chosen by trial and error for the different sequences and may change from one sequence to another. Large values for parameter α_3' will favor the smoothing of the velocity field, hence this parameter should be increased for noisy images. Typically $\alpha_3' = 200$ was used for synthetic sequences whereas α_3' was set to 1000 for real world sequences. Parameter α_3 acts like a threshold for detecting motion boundaries: small values for α_3 makes the process sensitive to small variations in the vector field and thus increases the number of detected motion boundaries whereas large values only retain the most relevant discontinuities. The final results were not very sensitive to the value of the α_3' smoothing parameter: a large range of values gave similar results. They appear more sensitive to parameter α_3 which has to be tuned accurately in order to get the relevant motion boundaries (like with a standard edge detector). Of course, an efficient data driven parameter identification method would be of great interest here.

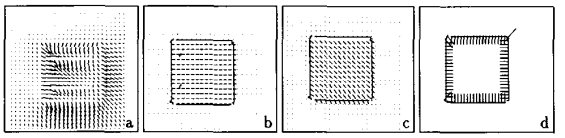


Fig. 7. Velocity fields and motion discontinuities obtained on the *square sequence*. (a) Optical flow estimation using a standard gradient-based smoothing method. (b) Optical flow and motion discontinuities obtained with our algorithm but without the moving edge constraint ($\alpha_2 = 0$). (c) Optical flow and motion discontinuities obtained by the complete multimodal estimation scheme. (d) Differenciation of occluding surface and occluded surface. Each symbolic vector points to the interior of the occluding region.

In our experiments, between 200 and 400 iterations (i.e. full scans of the image) are usually necessary to lead to configurations close to convergence. This number of iterations seems unusually large for a deterministic scheme. This is due to the fact that the velocity vectors are continuous-valued: in discrete problems, convergence is generally faster (within 10 iterations for binary-valued fields, for example). The number of iterations can be reduced by using a multiresolution implementation and other extensions described in Section V.

D. Comments about the Different Elements of the Model

The contributions of the different elements of the multimodal model are highlighted on the "square sequence" (Fig. 1(a)–(b)). Let us recall that, for this sequence, the grey value function *remains constant* in the square along the vertical direction and shows *a linear variation* along the horizontal one. Two bright horizontal lines (creating sharp intensity discontinuities) have been added inside the square (Fig. 1(a)). This sequence is typical of indoor scenes with texture-free objects which are difficult to handle using standard smoothing algorithms.

Fig. 7(a) shows again the flow measurements obtained by considering only the image flow constraint (taking into account neither the moving edge constraint, nor motion discontinuities nor confidence factors). This field is of the kind of what can be obtained with the standard Horn and Schunck's algorithm, [18]: the vector field is strongly corrupted near occlusions and intensity discontinuities and it appears oversmoothed.

Fig. 7(b) presents the result of multimodal estimation when motion discontinuities and the confidence factor ξ_g are used in the estimation process, but the moving edge constraint is discarded. The confidence factor ξ_g (see Fig. 3c) validates locally the image flow constraint (1). This constraint is invalidated on the boundaries of the square and on the two horizontal lines inside the square. Although the motion boundaries are precisely detected in this case, one can verify that only the horizontal component of the velocity field can be recovered using the image flow constraint, since the grey value function in the square remains constant along the vertical direction. More information is required to recover the second component of velocity. This information can be obtained from the two horizontal lines inside the square and from the square boundaries via the moving edge constraint.

Fig. 7(c) shows the result considering the full multimodal model obtained by adding the moving edge constraint and confidence factors ξ_{me} (see Fig. 3(b)). The moving edge constraint (2) yields reliable measurements for the points corresponding to intensity discontinuities, i.e. the boundaries of the square and the two horizontal lines inside the square. The full model (Fig. 7(c)) actually captures the motion boundaries and excellent accuracy is reached, even in the vicinity of discontinuities as can be seen. In this example the estimation process propagates the information gained from the moving edge constraint from the two central intensity lines and the square boundaries to the rest of the field. The cooperation between the moving edge constraint and the image flow constraint allows recovery of the horizontal and vertical components of the velocities. The sign of γ_d (Fig. 5) expresses, for every motion discontinuity, the relative position of occluding and occluded surfaces (Fig. 7(d)). Each symbolic dash in Fig. 7(d) points to the interior of the occluding region. The result is the right one, except for one point on the boundary.

This example suggests that, when the intensity profile shows limited variations, (creating ambiguous motion information in some areas) a cooperation between different motion cues is necessary to recover a complete motion vector. The usefulness of determining motion discontinuities appears also clearly. Finally, the confidence factors, associated to each motion constraint avoid to introduce inconsistent constraints, when the underlying models are broken.

The relative importance of the different elements of the model will typically depend on the class of images which are considered. In the case of natural outdoor images, with texture, the image flow constraint might be sufficient to recover significant motion information. The motion discontinuities and validation factor should of course be used to avoid smoothing near motion boundaries and to prevent a misuse of the image flow constraint in occlusion areas for instance. In the case of indoor scenes, with little texture, the contribution of moving edges can become relevant to recover the motion field in ambiguous regions.

IV. Experiments

Experiments have been carried out on several synthetic and real world sequences, involving qualitative and quantitative evaluation of the performance of the method. Indeed, we are usually able to judge the qualitative correctness of the estimated velocity fields, especially in the vicinity of motion discontinuities and occlusion areas.

When the true motion is known (this is the case in particular for synthetic sequences) a quantitative evaluation of the correctness of the field is possible by computing the difference between the estimated field and the true one and taking some norm of the difference field. Error histograms are also presented.

When the true motion is unknown (which is usually the case), one can consider frame-to-frame registration and compute some norm on the error between the motion compensated frame and the original one[1]. The root mean square error (RMSE) of the resulting difference image is computed in this

[1] More precisely, frame at time t is reconstructed from frame at time $t+1$ and from the estimated velocity field between t and $t+1$

case. It should however be noticed that RMSE only gives a very limited insight in the real correctness of a computed field. A perfect intensity image reconstruction does not necessarily mean that the computed motion field is physically consistent. Pel-recursive methods, though providing quite "surprising" motion fields, yet perform good motion compensation. Hence RMSE is not really an adequate evaluation of motion field correctness (but it remains the only available when the ground-truth is unknown !). For instance, in constant grey level areas RMSE does not depend on the computed velocity field. Moreover, a small error in the vicinity of a sharp intensity edge will have a worse effect on RMSE than a large error in an area of slowly varying intensity. We have also noticed that, even if the frame is compensated by the true displacement field, the RMSE can remain arbitrary large if large occlusion areas exist in the scene. Occlusion areas can of course not be compensated properly by a frame-to-frame registration, since points belonging to them have no correspondence in the other frame.

We have focused here on five sequences corresponding to different classes of images and movements: indoor scenes and outdoor scenes, situations comprising static camera and moving objects and situations involving both camera and object motions. These sequences have been chosen to illustrate different contributions of the multimodal estimation process: preservation of motion boundaries, multimodal cooperation between different measurement sources, processing of occlusions...

A. Experiments on Synthetic Sequences

The contributions of the different modeling parts are best highlighted on a synthetic sequence, in which motion can be controlled. The *moving square sequence* example (Fig. 7) has already shown that, even for the very ambiguous intensity pattern considered in that case, the multimodal motion estimation scheme recovers an optical flow field close to the true motion.

The errors between the true motion and the estimated fields presented in Fig. 7(a), (b), and (c) have been evaluated using the following norms: $L_1 = \frac{1}{N} \sum_{s \in S} \| \vec{\omega}_{estim}(s) - \vec{\omega}_{true}(s) \|$ and $L_2 = \left[\frac{1}{N} \sum_{s \in S} \| \vec{\omega}_{estim}(s) - \vec{\omega}_{true}(s) \|^2 \right]^{\frac{1}{2}}$, where N is the number of image sites.

The histogram of the local errors $\| \vec{\omega}_{estim}(s) - \vec{\omega}_{true}(s) \|$, $s \in S$, which gives information on the distribution of errors has also been computed (Fig. 8).

Quantitative evaluations of the results are presented in Table I. They make apparent the large errors in the flow field when certain elements of the multimodal scheme are discarded. The multimodal method divides the L_2 error by 4, with respect to the standard smoothing method.

The error histograms (Fig. 8) show that, in the case of standard smoothing (Fig. 7(a)), the errors are large and spread over a large range of values. For the full multimodal method (Fig. 7(c)) the error is strongly reduced and mainly concentrates near zero.

We present a second synthetic sequence called *moving disks*, including two disks undergoing different motions and occluding each other (Fig. 9). The foreground disk moves

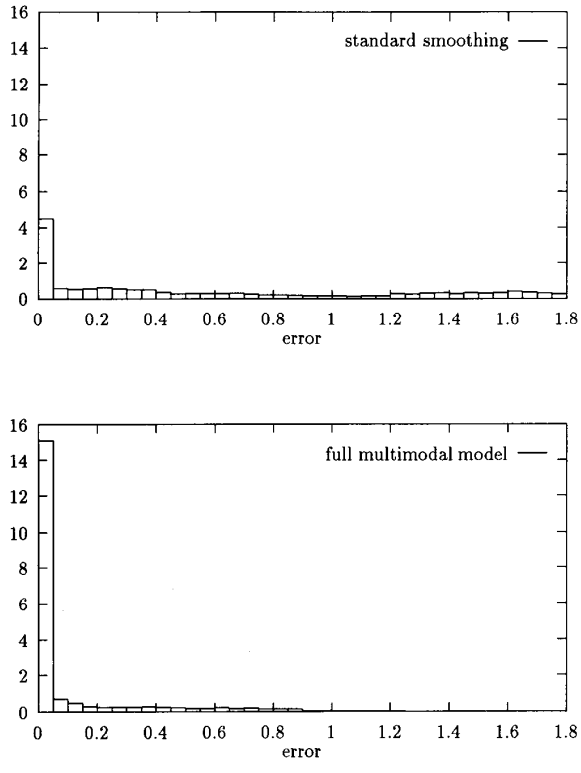


Fig. 8. Histogram of the error vector field on the *square sequence*.

TABLE I
ERRORS ON THE SQUARE SEQUENCE. THE ORIGINAL SEQUENCE IS PRESENTED IN FIG. 1. THE DIFFERENT METHODS ARE DESCRIBED IN SECTION III-D

METHOD	L1 error	L2 error
standard smoothing (see Fig. 7a)	1.01	1.42
without moving edge constraint (see Fig. 7b)	0.35	0.83
multimodal model (see Fig. 7c)	0.13	0.37

parallel to the axis of view, hence a dilatation is observed. This disk partially occludes another disk undergoing a translation of $3\sqrt{2}$ pixels toward the lower right corner. White noise (with variance 4.) has been added to the background.

The validation factors concerning the gradient-based motion constraint are derived from hypothesis test (7). Sites with validation factors equal to 0 are shown in Fig. 9(b). One can notice that they are closely related to the different occlusion areas in the scene: part of the background covered by the two moving objects and part corresponding to the overlapping of the two disks.

Fig. 9(c) depicts both the intensity edges detected on the original grey-level image and the perpendicular velocity components obtained from the ME estimator (see Section II-B). The intensity edges are somewhat noisy, due to white noise added to the background. The velocity components in Fig. 9c measure, as expected, the displacement of the moving disks perpendicularly to the intensity edges.

The fields computed after the two-step relaxation process are drawn in Fig. 9(d) and 9(e). 183 iterations were nec-

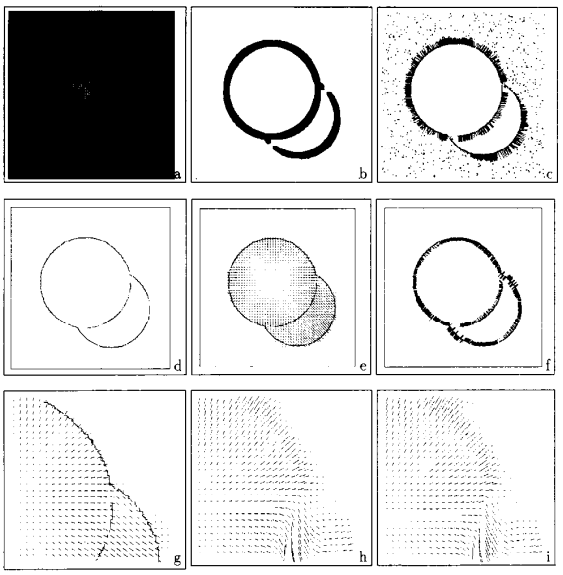


Fig. 9. Motion estimation on the *moving disks sequence*. (a) First frame from the original sequence (256 x 256). The foreground disk undergoes a dilatation, the background disk a translation. White noise has been added to the background. (b) Confidence factor for the gradient-based equation. The black areas corresponding to $\xi_g(s) = 0$ show the different occlusions in the scene. (c) Intensity edges detected on the first frame by Deriche's edge detector, [10]. The perpendicular velocity displacements computed by the moving edge estimator, have been superimposed on the intensity edges. (d) Motion boundaries estimated by the multimodal scheme (corresponding to $|\gamma_d(s)| = 1$). (e) Resulting optical flow field (183 iterations, $\alpha_3' = 200, \alpha_3 = 0.1$) (f) In this figure, each dash, computed from the sign of γ_d, points to the inner part of the occluding region. The disks are occluding the background. The occluded regions correspond to the background and the overlapping area between the two moving disks. (g) Upper right part of Fig. 9(e)) (100 x 100). (h) Result of Horn and Schunck's method. This field can be compared with the result of the multimodal scheme (Fig. 9(g)). (i) Result obtained without taking the occlusions into account but handling the motion boundaries. This shows that a major part of the estimation error comes from the occlusion areas.

essary here to reach convergence (model parameters: $\alpha_3' = 200, \alpha_3 = 0.1$). Fig. 9(d) presents the motion discontinuities corresponding to $|\gamma_d| = 1$. The proposed Markov interaction model filters the noisy detections on the background and only captures the true motion boundaries. Moreover, the sign of the motion discontinuity labels γ_d allows us to differentiate the occluding regions from the occluded ones. In Fig. 9(f), each dash points to the inner part of the occluding region. The performance of the method in the overlapping area between the two moving objects should be noticed: the dashes point to the true occluding disk. The result is the right one, except for two small parts of the boundary of the second disk which in fact slide parallel to themselves. There the local information remains ambiguous and it is not possible to differentiate the occluding region from the occluded one. Fig. 9(e) contains the estimated motion field: the accuracy is very satisfactory, when compared with the theoretical values, especially near the motion boundaries.

Again, these results have been compared to standard smoothing obtained by discarding motion discontinuities, occlusion areas and the moving edge constraint. Details of the fields computed using different methods appear in Fig. 9(g), (h), and (i). Fig. 9(g) corresponds to the upper right part of the field estimated by our multimodal method (Fig. 9(c)). Fig. 9(h) presents the results of the standard smoothing method. As expected, the resulting motion field is blurred across the motion discontinuities. *However* introducing motion discontinuities without processing occlusions is not sufficient, as shown in Fig. 9(i). The field in Fig. 9(i) is computed while handling local motion boundaries, but without considering the invalid sites of Fig. 9(b). Let us recall that these sites correspond to the occlusion areas. As a result, the final field is also very corrupted in this region. By comparing Fig. 9(h) and (i) one can conclude here that the major error source comes from the occlusion area rather than from an oversmoothing of the velocity field. This demonstrates that a specific processing of regions containing invalid observations is a real contribution of the multimodal scheme. Error statistics have been computed in this case for standard smoothing and multimodal estimation (II). The multimodal method brings significant improvement in field accuracy: a factor of 2 is obtained in this case with respect to standard smoothing.

TABLE II
ERRORS ON THE MOVING DISKS SEQUENCE.

METHOD	L1 error	L2 error
standard smoothing	0.58	1.04
multimodal model	0.26	0.48

B. Experiments on Real World Sequences

Several real-world sequences have been processed, which can be related to different contexts: traffic scene, TV sequences, etc. Quantitative results are presented (other experiments on real-world sequences may be found in [16]).

A first sequence called *interview* consists of a TV sequence: the woman on the right moves up and the camera follows her motion (Fig. 10).

A second sequence *houses* has been acquired by panning an urban scene. Prominent grey level features appear on the houses along with quite uniform regions (Fig. 11).

The results presented here are computed from two successive frames out of the original sequences.

As far as the *interview* sequence is concerned, motion boundaries are closely related to the woman's movement (compare Fig. 10(b) with Fig. 10(c)). The estimated motion field accurately reproduces the visual motion in the background due to the camera panning. The complex motion of the woman moving up (Fig. 10(d)) is recovered with good accuracy, especially near motion boundaries. It can be compared to the field computed by the standard smoothing method, Fig. 10(f). The oversmoothing is very perceptible in this last case. The root mean square error are respectively 9.12 (multimodal scheme) and 9.37 (standard smoothing). The difference between the two methods seems small because the standard smoothing method does a good job in this example nearly everywhere, but in the motion boundary areas whose size is negligible compared to the image size. Again, one must make use of this global quality measure with caution; it cannot account for local artefacts which nevertheless are detrimental.

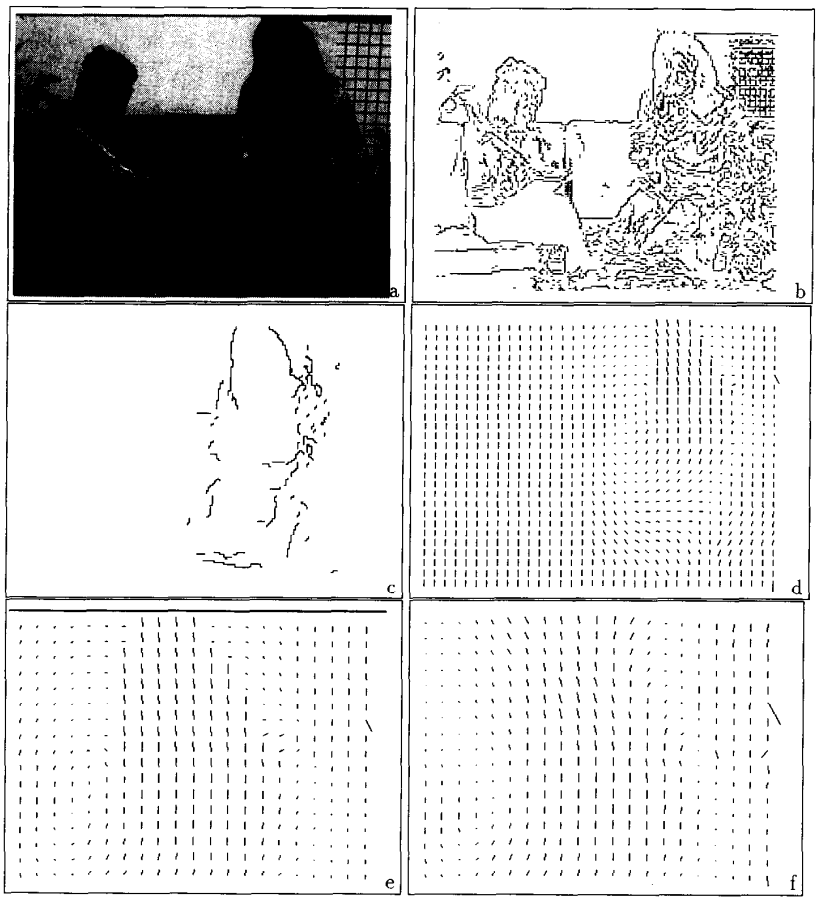


Fig. 10. Multimodal motion estimation: *Interview sequence* (by courtesy of BBC-UK). (a) First frame of the original sequence (134 x 168): the woman at the right moves up and the camera follows her motion. (b) Intensity edges extracted from Fig. 10(a). (c) Motion boundaries estimated by the multimodal estimation scheme (400 iterations,$\alpha_3' = 1000..\alpha_3 = 0.15$). (d) Associated optical flow estimation (horizontally and vertically subsampled by 3). (e) Detail of the optical flow field of Fig. 10(d) showing the woman's head. (f) Result of standard smoothing on the same detail as in Fig. 10(e). The smoothing of the optical flow field across the discontinuities is visible.

The sequence *houses* (Fig. 11) was chosen to show the contribution of the moving edge constraint, when there are many regions of uniform intensity in the image. This is the case in the almost uniform areas of the house roofs and walls in Fig. 11(a). Fig. 11(b) shows moving edges detected by the algorithm described in [7]. Fig. 11(c) and 11(d) present the resulting velocity field computed using two different methods. The first one only makes use of the gradient-based constraint, the second one includes both the gradient-based and the moving edge constraint. A low value was chosen here for the smoothing parameter α_3' ($\alpha_3' = 10$) in order to emphasize the differences between the two versions. The improvement due to the multimodal cooperation scheme is visible (Fig. 11(d)). The visual motion corresponding to a translation in the image plane is better estimated in the vicinity of moving edges in Fig. 11d than in Fig. 11c. The image flow equation here does not bring sufficient local information: the use of an additional constraint significantly improves the result. As far as the frame-to-frame registration is concerned, the root mean square of the error image is 14.0 in the first case and 9.4 in the second one.

V. Multiresolution Motion Analysis

As explained in Section II-A, large displacements are generally not reachable using gradient-based motion estimation methods. This also concerns our multimodal scheme, which makes use of the gradient-based constraint. As soon as the displacements become large, the confidence factor associated to the gradient-based measurements decreases very quickly. As a consequence, a large part of the gradient-based measurements do not take part to the estimation of the final velocity field and in many cases there is not enough information available to get reasonable results. Besides, iterative relaxation schemes are very slow to propagate velocity information into image areas with almost homogeneous or linearly sloping grey value distribution, [11]. A standard solution to these problems is to use a multiresolution image analysis, [3], [4],

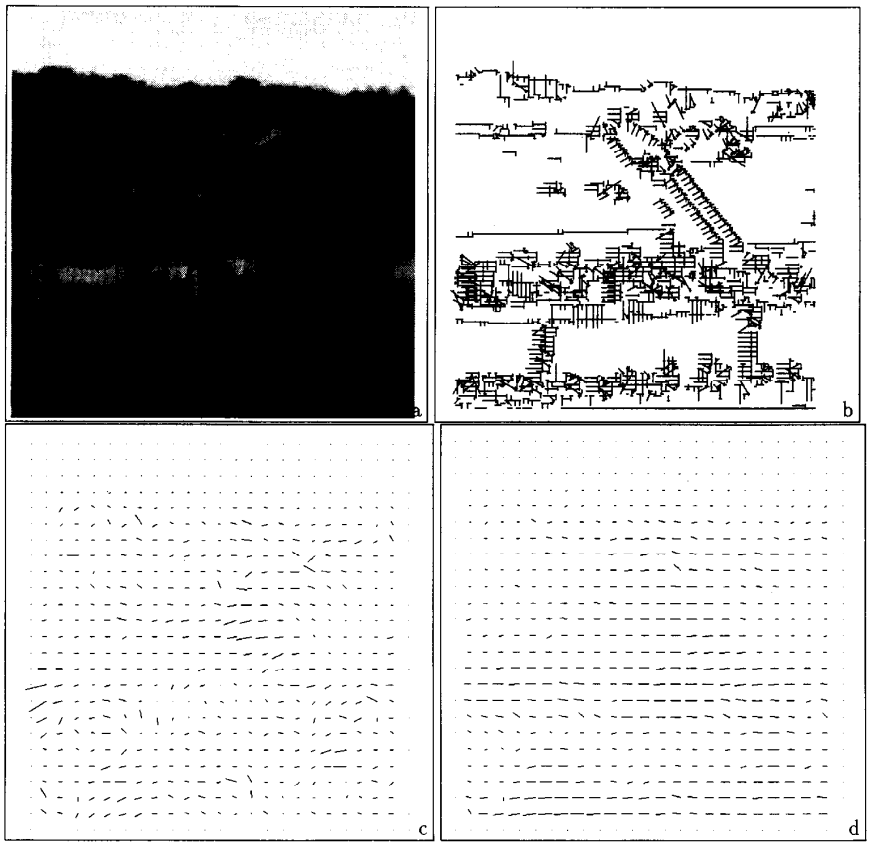


Fig. 11. Multimodal motion estimation: *houses*. (a) First frame of the sequence (170 x 170) .(b) Perpendicular velocity components estimated on the intensity edges. (c) Gradient-based only optical flow estimation (400 iterations, $\alpha_3{}' = 10.$, $\alpha_3 = 1.$). (d) Multimodal optical flow estimation (horizontally and vertically subsampled by 7) (same parameter values as in Fig. 11 (c)).

[11], [22], [33]. The large displacement vectors are determined at lower resolutions, where the interactions between spatial and temporal derivatives are maintained. The multimodal estimation scheme, leading to a standard relaxation algorithm, is naturally well suited to multiresolution processing.

Multigrid techniques (usual in numerical analysis) have been adapted to visual motion computation by Terzopoulos, [33] and Enkelmann, [11]. Multiresolution methods have been proposed for MRF-based relaxation algorithms by Barnard, [3] for stereo matching and Konrad *et al.*, [22] for motion estimation.

The implemented multiresolution algorithm consists of a coarse-to-fine strategy, starting from the lowest resolution and propagating the estimates from the coarse scales to the finer ones. A gaussian image pyramid is built up using low-pass filtering and subsampling by a factor of 2 the original images of the sequence. The optical flow at resolution level k is denoted $\vec{\omega}^k$. Three levels of resolution are used in our experiments ($k = 0, 1, 2$). The multiresolution algorithm can be described as follows :

1) Estimation of the optical flow at the lowest resolution level ($k = 2$) using the original multimodal scheme.
2) Repetition and bilinear interpolation of the vectors from the coarse level k to the finer level $k - 1$. The interpolation takes into account the location of the motion boundaries. The interpolated field is denoted $\vec{\omega}_0^{k-1}$.
3) Estimation of an *incremental* optical flow field $\vec{d\omega}^{k-1}$ at level $k - 1$ introducing a modified version of the image flow equation, [11] in the global energy function:

$$\vec{\nabla} f(s + \frac{\vec{\omega}_0^{k-1}}{\Delta t}, t + \Delta t).\vec{d\omega}^{k-1}(s) + f(s + \frac{\vec{\omega}_0^{k-1}}{\Delta t}, t + \Delta t) - f(s,t) = 0$$

The relaxation is performed until convergence at that level (the convergence criterion is the same as in the single resolution case). The final optical flow field at level $k - 1$ is: $\vec{\omega}^{k-1} = \vec{\omega}_0^{k-1} + \vec{d\omega}^{k-1}$. The motion boundaries are estimated using the same energy function as in the single resolution method.
4) If the current level is 0, stop; else $k := k - 1$, goto 2.

In the experiments carried out, the same parameters values are used for the potential functions at each level of the image pyramid.

The contribution of the multiresolution relaxation method is illustrated here on one real-world sequence: a TV sequence called *Mobi* comprising large displacements, several different

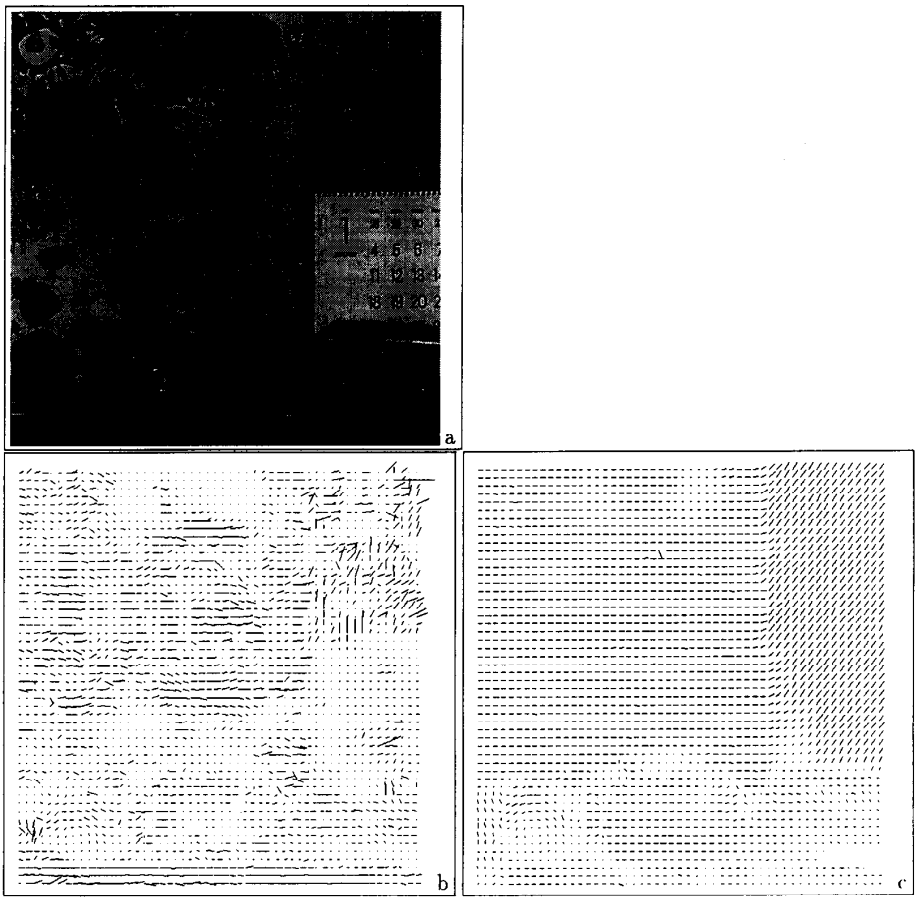


Fig. 12. Multiresolution estimation: *Mobi sequence* (by courtesy of CCETT-Rennes). (a) First frame of the original sequence (512 × 512): the scene is composed of a rolling ball, a moving toy-train and a calendar undergoing a vertical translation. The camera motion corresponds to a panning of the scene. (b) Optical flow estimation with the original single resolution scheme (horizontally and vertically subsampled by 10). (c) Optical flow estimation with the multiresolution method (horizontally and vertically subsampled by 10).

TABLE III
NUMBER OF ITERATIONS IN THE SINGLE AND MULTIRESOLUTION SCHEME (MOBI SEQUENCE).

	Level	Iteration No	Equ. Iter.
Multiresolution	0	43	56
	1	37	
	2	59	
Monoresolution	0	111	111

Equ. Iter.: Computational equivalent of one complete sweep through the image at the finest resolution.

moving objects and involving camera motion. The scene is composed of a rolling ball, a moving toy-train and a calendar undergoing a vertical translation (Fig. 12(a)). The camera motion corresponds to a panning of the scene, which yields an additional horizontal translation component in the optical flow.

For the "Mobi" sequence, due to large displacements along with important uniform areas (on the calendar and on the wallpaper for instance) and sharp edges, the spatial and temporal derivatives interact on a very short range in that case. Therefore, the final optical flow field computed by the original single resolution scheme is not satisfactory (see Fig. 12(b)). The estimates delivered by the multiresolution algorithm, with three resolution levels, are presented in Fig. 12(c). Visually the optical flow recovered in the multiresolution case is closer to the real underlying motion (see for example the apparent diagonal translation on the calendar). Table III shows the number of iterations required at each resolution level to reach convergence. The total iteration number in the multiresolution case corresponds to an equivalent number of 56 iterations at full resolution. This is half that of the single resolution algorithm which requires 111 iterations to converge to a result of lower quality. The improvement in frame-to-frame registration is highly significant here: RMSE is 27.8 for single resolution, 11.8 for multiresolution processing. Multiresolution processing brings a significant improvement in the quality of the estimated optical flow fields as well as an appreciable speed-up of the algorithm.

A second possible extension of the proposed scheme is related to the processing of multiple frames. Intrinsically, the method only considers two successive frames of the sequence. There are no connections between the motion estimates de-

rived at different times. However, the velocities usually vary smoothly along a sequence; hence, the estimated velocities at time t can be used to support the estimation at time $t + dt$. Such an extension is described in [16].

VI. Concluding Remarks

We have presented a general algorithm for optical flow estimation which is able to jointly handle discontinuities and occlusions in the motion field. It can be interpreted as a generalized regularization approach to the ill-posed problem of optical flow computation. The method has been called multimodal in that it integrates several complementary constraints on the desired solution. Statistical models express the interactions between the different low-level image entities: velocity vectors, motion boundaries, occluding and occluded surfaces, intensity edges and the spatio-temporal variations of the brightness pattern. The motion measurements are based on two complementary constraints: gradient-based and feature-based. The algorithm requires the tuning of only two main parameters which balance the smoothing of the velocity field and the sensitivity of the motion boundary detection. A multiresolution implementation of this algorithm has been described, which appears very efficient for the measurement of large displacements.

Experiments have been carried out on a large number of real-world sequences: outdoor and indoor scenes imaged by a moving camera with several moving objects and large displacements. One key feature of the described scheme is its ability to handle properly occlusion areas. Indeed, in image sequences, discontinuities are not only local but exist on large areas corresponding to occluded surfaces. This problem has been addressed here by testing directly the validity of the underlying motion measurement equations. We think this is an efficient way to cope with the general occlusion problem. The experimental results on synthetic sequences clearly demonstrate the advantages of this approach. We think that the multimodal estimation algorithm should be considered as a step toward a comprehensive multimodal motion estimation scheme. Such a scheme would enable velocity estimation in very general situations: for textured outdoor scenes as well as for structured man-made environments with long-range or short-range motion.

Yet, a perfect detection of motion boundaries remains difficult: for instance the extracted motion boundaries are sometimes locally broken. This is mainly due to two factors: the quality of intensity edges used as partial support for estimating motion boundaries and the use of a deterministic optimization algorithm which yields suboptimal motion edge configurations. However, the statistical framework described here seems flexible enough to allow several important extensions. For instance, in the present scheme, only local motion boundaries are determined. An extension toward a region-based motion segmentation algorithm would be of interest in the context of dynamic scene analysis. A partition of a sequence into its constituent moving objects indeed defines a first key-step in many dynamic scene analysis problems. Such an extension can be found in [8].

Another straightforward addition could be to introduce in the multimodal cooperation process other local motion measurements resulting from similarity functions, token tracking or grey-value corners matching, for instance.

Besides, a significant contribution to MRF modeling would be the development of a consistent and tractable theoretical framework for multiresolution MRF-based image analysis. A recent contribution to this problem may be found in [17].

The last point is the temporal stability of the extracted motion cues (motion boundaries or regions). In the algorithm described here, there is no *a priori* modeling of the connections which naturally exist between estimates obtained at different times. It would be of interest to introduce also a control on the temporal dimension, in order for example to filter and to track motion cues along the sequence. A class of models involving temporal neighborhoods has already been introduced in motion detection [9], motion segmentation, [25] and in motion measurement, [6]. They appear promising as far as the processing of long sequences is concerned.

Acknowledgment

The authors would like to thank T. Catudal for the implementation of the multiresolution method discussed in Section V.

References

[1] G. Adiv, "Determining three-dimensional motion and structure from optical flow generated by several moving objects," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 7, pp. 384–401, July 1985.

[2] J. K. Aggarwal and N. Nandhakumar, "On the computation of motion from sequences of images—A review," *Proc. IEEE*, vol. 76, no. 8, pp. 917–935, 1988.

[3] S. T. Barnard, "Stochastic stereo matching over scale," *Int. J. Comput. Vision*, vol. 3, pp. 17–32, 1989.

[4] J. Bergen, P. Burt, R. Hingorani, and S. Peleg, "Computing two motions from three frames," in *Proc. 3rd Int. Conf. Comput. Vision*, Osaka, Dec. 1990, pp. 27–32.

[5] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc.*, vol. 48, ser. B, no 3, pp. 259–302, 1986.

[6] M. J. Black and P. Anandan, "A model for the detection of motion over time," in *Proc. 3rd Int. Conf. Comput. Vision*, Osaka, Dec. 1990, pp. 33–37.

[7] P. Bouthemy, "A maximum-likelihood framework for determining moving edges," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 5, pp. 499–511, May 1989.

[8] P. Bouthemy and E. François, "Motion segmentation and qualitative dynamic scene analysis from an image sequence," *Int. J. Comput. Vision*, vol. 10, no. 2, pp. 157–182, 1993.

[9] 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 Conf. Comput. Vision*, Antibes, France, Apr. 1990, pp. 307–311.

[10] R. Deriche, "Using Canny's criteria to derive a recursively implemented optimal edge detector," *Int. J. Comput. Vision*, pp. 167–187, 1987.

[11] W. Enkelmann, "Investigations of multigrid algorithms for the estimation of optical flow fields in image sequences," *Comput. Vision, Graphics, Image Processing*, vol. 43, pp. 150–177, 1988.

[12] E. Gamble and T. Poggio, "Visual integration and detection of discontinuities: The key role of intensity edges," Tech. Rep. A.I. 970, M.I.T., Oct. 1987.

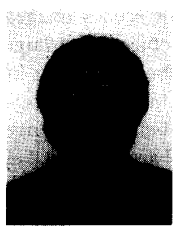
[13] D. Geman, S. Geman, C. Graffigne, and D. Pong, "Boundary detection by constrained optimization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 7, pp. 609–628, July 1990.

[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, pp. 721–741, Nov. 1984.

[15] F. Heitz and P. Bouthemy, "Multimodal motion estimation and segmentation using Markov random fields," in *Proc. 10th Int. Conf. Pattern Recognit.*, vol. 1, Atlantic City, NJ, June 1990, pp. 378–383.
[16] ——, "Multimodal estimation of discontinuous optical flow using Markov random fields," Tech. Rep. 1367, INRIA-Rennes, Jan. 1991.
[17] F. Heitz, P. Perez, and P. Bouthemy, "Multiscale minimization of global energy functions in some visual recovery problems," *CVGIP: Image Understanding*, accepted for publication, 1993.
[18] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intell.*, vol. 17, pp. 185–203, 1981.
[19] J. Hutchinson, C. Koch, J. Luo, and C. Mead, "Computing motion using analog and binary resistive networks," *Comput.*, vol. 21, pp. 52–63, Mar. 1988.
[20] J. K. Kearney, W. B. Thompson, and D. L. Boley, "Optical flow estimation: An error analysis of gradient-based methods with local optimization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, no. 2, pp. 229–244, 1987.
[21] J. Konrad and E. Dubois, "Estimation of image motion fields: Bayesian formulation and stochastic solution," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, New York, 1988, pp. 1072–1075.
[22] ——, "Bayesian estimation of motion vector fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 9, pp. 910–927, 1992.
[23] J. J. Little and W. E. Gillett, "Direct evidence for occlusion in stereo and motion," in *Proc. First European Conf. Comput. Vision*, Antibes, France, Apr. 1990, pp. 336–340.
[24] A. Mitiche, Y. F. Wang, and J. K. Aggarwal, "Experiments in computing optical flow with the gradient-based, multiconstraint method," *Pattern Recognit.*, vol. 20, no. 2, pp. 173–179, 1987.
[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, pp. 220–228, Mar. 1987.
[26] H. H. Nagel, "On the estimation of optical flow: Relations between different approaches and some new result," *Artificial Intell.*, vol. 33, p. 299–324, 1987.
[27] H. H. Nagel and W. Enkelmann, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, pp. 565–593, 1986.
[28] S. Peleg and H. Rom, "Motion based segmentation," in *Proc. 10th Int. Conf. Pattern Recognit.*, vol. 1, Atlantic City, NJ, June 1990, pp. 109–113.
[29] B. G. Schunck, "Image flow segmentation and estimation by constraint line clustering," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 10, pp. 1010–1027, Oct. 1989.
[30] M. Shizawa and K. Mase, "Simultaneous multiple optical flow estimation," in *Proc. 10th Int. Conf. Pattern Recognit.*, vol. 1, Atlantic City, NJ, June 1990, pp. 274–278.
[31] A. Singh, "An estimation-theoretic framework for image flow computation," in *Proc. 3rd Int. Conf. Comput. Vision*, Osaka, Dec. 1990, pp. 168–177.
[32] A. Spoerri and S. Ullman, "The early detection of motion boundaries," in *Proc. First Int. Conf. Comput. Vision*, London, U.K., June 1987, pp. 209–218.
[33] D. Terzopoulos, "Image analysis using multigrid relaxation methods," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, no. 2, pp. 129–139, Mar. 1986.
[34] A. Verri and T. Poggio, "Motion field and optical flow: Qualitative properties," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 5, pp. 490–498, May 1989.
[35] K. Wohn and A. M. Waxman, "The analytic structure of image flows: deformation and segmentation," *Comput. Vision, Graphics, Image Processing*, vol. 49, pp. 127–151, 1990.

Fabrice Heitz was born in France in 1961. He graduated from "Telecom Bretagne," France, in 1984 and received the Ph.D. degree from "Telecom Paris," France, in 1988.

Since 1988, he has been with INRIA Rennes as a full-time researcher in Computer Vision. Previously, he has worked at ENST and LRMF (Laboratoire de Recherches des Musées de France-Louvre Museum), Paris, on Fine Arts Analysis using Image Processing techniques. His current fields of interest are statistical image modeling, multiscale image processing, dynamic scene analysis, deformable models, and parallel algorithms and architectures.



Patrick Bouthemy was born in France in 1957. He graduated from Ecole Nationale Supérieure des Télécommunications, Paris, in 1980, and received the Ph.D. degree in computer science from the University of Rennes in 1982.

From December 1982 until February 1984, he was employed by INRS-Télécommunications, Montréal, PQ, Canada, in the Department of Visual Communications. Since April 1984, he has been with INRIA at IRISA in Rennes. His major research interests are motion analysis, statistical models for image sequence processing, qualitative vision, and active vision.