# VISUAL MODULE INTEGRATION FOR OPTICAL FLOW ESTIMATION

*Luigi Bedini†, Andrea Cannata†, Mario Ferraro‡, Emanuele Salerno†, Anna Tonazzini†*
†Istituto di Elaborazione della Informazione - CNR,
Via Santa Maria, 46, 56126 Pisa, ITALY
‡Dipartimento di Fisica Sperimentale, Università di Torino,
Via Giuria, 1, 10125 Torino, ITALY
e-mail: {bedini,salerno,tonazzini}@iei.pi.cnr.it, ferraro@ph.unito.it

**ABSTRACT**

A technique to integrate gradient-based and feature-based modules to estimate the optical flow from a pair of images is proposed. The integration strategy is based on a Bayesian approach, where the optical flow is evaluated as the minimizer of a suitable posterior energy function, containing all the gradient and feature information on the problem. The capability of the technique to constrain the displacement in the neighbourhoods of motion discontinuities has been tested.

## 1 INTRODUCTION

Estimating the optical flow from an image sequence is a very important task for all motion-based visual functions [1].

The basic problem that must be solved in determining optical flow is to register a pair of successive frames of the sequence available. Two basic classes of approaches, gradient-based and feature-based, have been taken into consideration to solve this problem. The gradient-based approaches are based on the assumptions of uniform illumination, Lambertian surfaces, rigid motion and small displacements, and look for a match between the two frames by taking into consideration the gradients of the image intensity function. The visual cues used for registration are pixel intensities. The output produced by these methods is a 'dense' (defined for each image pixel) map of displacement vectors. Image registration from intensity gradients is an ill-posed problem, normally regularised by introducing smoothness constraints. As in other visual problems, this is not sufficient to reach a good solution, in that smoothing is not physically justified where discontinuities are present. In this particular case, moreover, there are additional difficulties in handling the occlusion zones, where the displacement vector is not even defined. The presence of these zones and the smoothness constraint cause significant errors in optical flow estimation, in that the errors are propagated throughout the image. It is thus clear that this problem should be further regularised. The additional information to be introduced could be some piece of specific prior knowledge, additional sensor data, or complementary information coming from other computational modules. As an early attempt, the use of discontinuities has been proposed [2]. In [3] a gradient-based method regularised by edge-preserving local smoothness is proposed, where displacement discontinuities are detected with the help of intensity discontinuities.

In the feature-based approaches, the visual cues used for the matching procedure are high level features, such as particular geometric patterns, which can be detected in the different frames with a certain confidence. For the case of features such as corners, the hypotheses on which these methods are based are that any single cue extracted from one image corresponds to the same physical point that is projected in the corresponding cue in the other image. This means that, if the features are carefully selected, the displacements evaluated are highly reliable. On the other hand, the displacement map obtained from a feature-based method is 'sparse' (it is not defined for all the image pixels).

A way to overcome the drawbacks of both purely gradient-based and purely feature-based methods, thus producing dense and reliable optical flow maps, could be to integrate gradient-based and feature-based approaches. For instance, this can be done by including the feature-based information into a suitable energy function. An approach that explicitly addresses this principle is proposed in [4], where the computational network is formed by a feature-based module and a gradient-based module, which also uses prior information on local smoothness. The geometric features taken into account in that case are the intensity discontinuities contained in both the data images, which are detected and matched by means of a local analysis, under the assumption of very small displacements. This feature-based module, however, is only capable of estimating the orthogonal motion component, because of the well-known aperture problem, and this causes troubles where edge motion has no orthogonal component.

To better constrain the motion field, we propose herein to replace the orthogonal motion detector by a computational module that estimates the complete displacement vector for entire rectilinear segments. This

can be made by extracting segments whose ends can be reliably identified and matched in both images, thus overcoming the aperture problem. The displacement vectors thus calculated can be integrated in the energy function derived from the gradient-based constraints. An experimental evaluation of the better performance of this method, when compared to the one proposed in [4], is being carried out.

## 2 THE PROPOSED TECHNIQUE

### 2.1 Integration of gradients and features

The scheme adopted to integrate the different computational modules is shown in Figure 1. The features considered at present are rectilinear segments, whose ends are marked by highly reliable points, such as corners, crosses or junctions. The first step to identify the useful features is to extract line drawings from the gray level or color images available. Since we limit our interest to straight lines, this can be done by edge extraction followed by a procedure that isolates the straight-line segments from the edge map [5, 6, 7]. From the line drawing we then extract, by graph search, the patterns we are interested in. Finally, we have to match corresponding patterns in the two frames. This can be done by means of subgraph isomorphism [8, 9]. The validity of our method is based on the hypothesis that any line pattern extracted from one image does not change its topology in the other image, and is related to a rigid object. In this case, given the ends of a segment in one image and the corresponding points in the other image, we can easily evaluate the two components of the displacement at these points and (by simple interpolation) at each point in the segment. This process is performed by the block "Feature-based motion evaluation" in Figure 1, which receives as its input the intensity edges, extracted by an independent module. The values of the displacements thus obtained will be used to constrain the optical flow in the rectilinear moving edge zones, by introducing them in the final energy function.

The "Gradient-based motion evaluation" module extracts the spatial and temporal derivatives from the input sequence. It has been shown [4] that the gradients and the geometric features are complementary data. This means that the related constraints are valid over separate zones of the flow field, namely, the smooth zones for the gradients and the discontinuities and occlusion zones for the geometric features. The selection between these constraints is performed by the switch shown in the figure, which is driven by two binary validation factors, calculated as described in [10].

The fusion of the gradient- and feature-based information is performed by the two interacting blocks "optical flow" and "Motion discontinuities". The flow field is modelled as a pair of interacting Markov Random Fields: one is a continuous vector field representing the displacements, and the other is a binary field representing motion discontinuities. The estimation is also supported by prior knowledge, namely, the smoothness constraint and some known geometric properties of the discontinuity set, enforced by means of specially designed MRF potentials (see 2.2 below). In particular, some unlikely configurations of the lines, such as double lines, crosses and T-junctions, are discouraged in that their appearance is rare in the images [11]. The intensity edges help detecting the motion edges in that, as pointed out in [2], motion discontinuities are likely to appear in a subset of intensity discontinuities. For this reason, we strongly discourage the formation of motion edges where no intensity edge is detected.

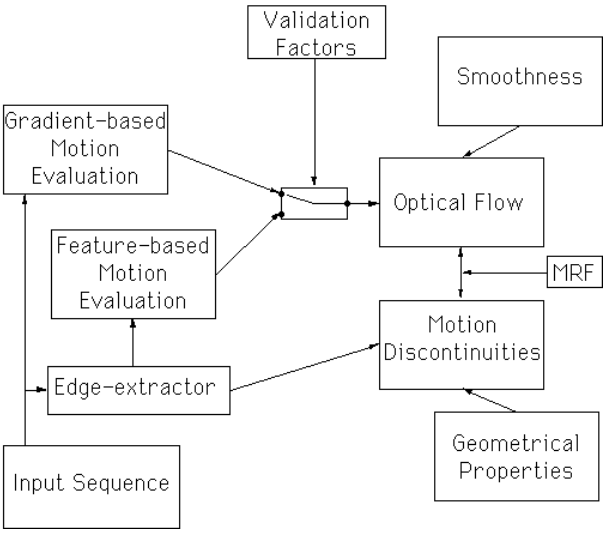


Figure 1: Computational network adopted

### 2.2 The posterior energy

The energy function is built on the basis of Bayesian integration among the various modules shown in Figure 1. In a discrete setting, the final form of our criterion is:

$$E(\omega,\gamma|f,\xi_g,\xi_{me},\eta,w) =$$

$$\alpha_1 \sum_{s\in S} \xi_g(s)(f(s;t+\delta t) - f(s+\omega_s;t))^2$$

$$+\alpha_2 \sum_{(t,d)\in\{C_1\}\cup\{C_2\}} \xi_{me}(d)||\omega_t - w_d||^2 \frac{1}{2}(-|\gamma_d| + \gamma_d + 2)+$$

$$+\alpha_2 \sum_{(s,d)\in\{C_3\}\cup\{C_4\}} \xi_{me}(d)||\omega_s - w_d||^2 \frac{1}{2}(-|\gamma_d| - \gamma_d + 2)+$$

$$+\alpha_3 \sum_{(s,t,d)\in\{C_5\}\cup\{C_6\}} (||\omega_s - \omega_t||^2 - \beta^2)(1 - |\gamma_d|)+$$

$$+\alpha_4 \sum_{d \in D} (1-\eta_d)|\gamma_d| + \sum_{c \in \{C_7\} \cup \{C_8\}} V_c(\gamma) \qquad (1)$$

where: $\omega$ is the matrix of the unknown displacements; $\gamma$ is the matrix of the three-valued $(+1;0;-1)$ parameters indicating the displacement discontinuities; $f(s;t)$ is the data image sequence; $\xi_g$ and $\xi_{me}$ are binary validation factors that mark the regular zones and the moving edges, respectively [4,10]; $\eta$ is the matrix of the binary $(+1;\ 0)$ intensity discontinuities, defined on the same site set $D$ as $\gamma$, and evaluated in advance by the module "edge-extractor" in Figure 1; $w$ is the moving-edge displacement matrix, as calculated by the feature-based procedure; parameters $\alpha_{1-4}$ are regularisation coefficients, and $\beta$ is a threshold for the creation of a discontinuity. The summation weighted by $\alpha_1$ is the data consistency term. The terms weighted by $\alpha_2$ introduce the feature-based information. The term with $\alpha_3$ enforces smoothness where there is no displacement discontinuity, and the term with $\alpha_4$ penalizes displacement discontinuities that do not occur in intensity discontinuities (see [2]). Finally, the last summation enforces geometric constraints on the displacement discontinuity curves; the functions $V_c(\gamma)$ are suitable potentials associated to cliques $c$ of displacement discontinuities. The summations are performed over the clique sets shown in Figure 2, associated with the displacement site set $S$ and the discontinuity site set $D$. A short explanation is due on the meaning of the three values allowed for the variables $\gamma$. As is seen from the summations weighted by $\alpha_2$, if the absolute value of $\gamma_d$ is 1, then only one of these summations takes a positive contribution from the edge element $d$. More specifically (see again Figure 2), if $\gamma_d = -1$, the pixel $s$ will be constrained to follow the motion of $d$, whereas the pixel $t$ belongs to an occluded zone, and no motion smoothness is enforced between $t$ and $d$. The converse is true if $\gamma_d = +1$: the pixel $t$ belongs to an occluding zone and the pixel $s$ belongs to an occluded zone. If $\gamma_d = 0$, both pixels $s$ and $t$ will follow the motion, $w_d$, of the edge element. This means that no motion discontinuity is present across $d$.

The energy function contains all the information, measured, evaluated, or available *a priori* on the problem, and realizes the integration among data fit, edge extraction, graph matching, and prior information. The difference between (1) and the posterior energy adopted in [4] is that in this case the total displacement vector is constrained onto the moving straight-line segments, whereas in [4] each moving edge was constrained for only the normal motion component.

This is just one possible attempt to integrate high-level and low-level modules in solving a visual problem. A more general approach could be envisaged by considering other features besides rectilinear segments. As an example, suitable invariant measures on particular geometric configurations could be exploited. This problem is presently being studied.

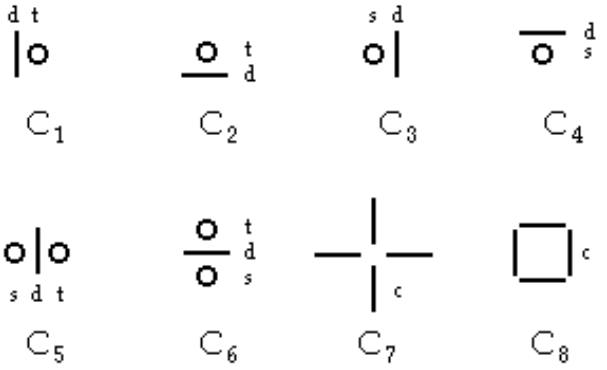


Figure 2: Clique system

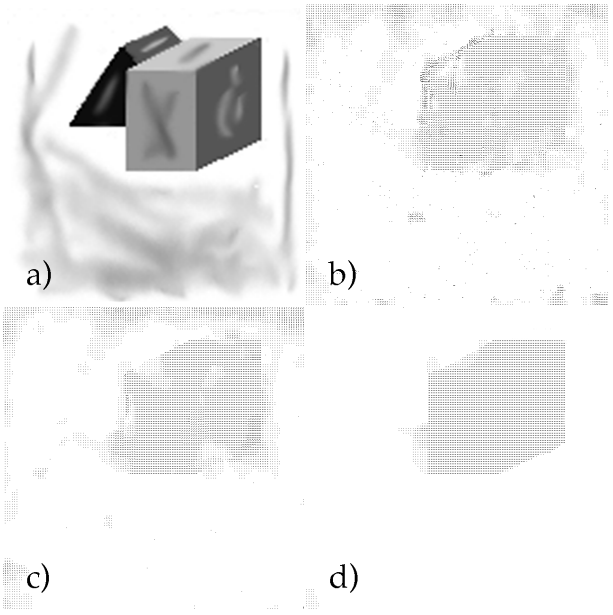


Figure 3: a) First input frame; b) result with no moving edge constraint; c) result form minimization of (1); best result from minimization of (1)

## 3 RESULTS

At present, the main goal of the experimental investigation is to evaluate the possible improvements obtainable by using the complete line displacement instead of its only normal component. In this section, we just show some preliminary results. To minimize the energy in $\omega$ and $\gamma$, we adopted a simulated annealing algorithm, with no concern with computational complexity, with the aim at best approaching the global energy minimum.

To impose the constraint on orthogonal motion over straight-line segments moving along their own directions is equivalent to imposing no constraint. A comparison between the two approaches is exemplified in Figure 3.

The first frame of the synthetic input sequence is shown in figure 3a). The only moving object in the scene is the cube, which translates leftwards; its lower edge thus translates along its own direction. To obtain the result shown in Figure 3 b), no constraint was placed on this line, and the optical flow there evaluated is affected by considerable errors. In Figure 3 c) we show the result obtained by constraining the total displacement along that line. We can easily note that in this case the optical flow on the lower edge of the cube is correctly evaluated, thus demonstrating the validity of our constraint. Figures 3 b) and 3 c) were obtained by using the same values for all the parameters involved in the calculation, with no particular attention at optimizing their values. In figure 3 d), we show the best result obtained by our method after a careful choice of both the free parameters and the annealing schedule.

## 4 CONCLUSIONS

A complete evaluation of the performance of the proposed technique cannot be performed at this point. Indeed, we do not know whether the advantages shown over the approach with the only normal component constraint are worth the possible complications in the algorithm. The zones where these advantages can be obtained are normally rather restricted in comparison with the entire solution field. A possibility to be taken into consideration is imposing both the total displacement and the normal component constraint. This can be useful, in that the total displacement can be calculated only where suitable features can be extracted, in our case some rectilinear segments, whereas the normal component constraint can be enforced everywhere the edge extractor finds an intensity line. These two constraints are thus not mutually exclusive, and can be both enforced in their respective areas of validity. Another way to improve the technique could be to identify other features to be considered, based as an example on translational invariants.

In parallel with these theoretical developments, the feature-based module should be integrated in the overall procedure, and a fast minimization algorithm should be selected, in order to be able to fully evaluate the performance of the method.

## 5 REFERENCES


1. A. Mitiche, P. Bouthemy, "Computation and Analysis of Image Motion: A Synopsis of Current Problems and Methods",*Int. J. of Computer Vision*, **19**, 29-55 (1996).

2. E. Gamble, T. Poggio, "Visual Integration and Detection of Discontinuities: The Key Role of Intensity Edges", Massachusetts Inst. Technol., *A.I. Memo No. 970* (1987).

3. J. Konrad, E. Dubois, "Bayesian Estimation of Motion Vector Fields", *IEEE Trans.*, **PAMI-14**, 910-927 (1992).

4. F. Heitz, P. Bouthemy, "Multimodal Estimation of Discontinuous Optical Flow Using Markov Random Fields", *IEEE Trans.*, **PAMI-15**, 1217-1232 (1993).

5. W.-N. Lie, Y.-C. Chen, "Robust Line-Drawing Extraction for Polyhedra Using Weighted Polarized Hough Transform", *Pattern Recognition*, **23**, 261-274 (1990)

6. W.K. Gu, T.S. Huang, "Connected Line-Drawing Extraction from a Perspective View of a Polyhedron",*IEEE Trans.*, **PAMI-7**, 422-430 (1985).

7. J. Burns, A. Hanson, E. Risemen, "Extracting Straight Lines", *IEEE Trans.*, **PAMI-8**, 425-455 (1986).

8. E.K. Wong,"Model Matching in Robot Vision by Subgraph Isomorphism", *Pattern Recognition*, **25**, 287-303 (1992).

9. G. Marola, A. Vaccarelli, "An Algebraic Method for Detection and Recognition of Polyhedral Objects from a Single Image", *Pattern Recognition*, **27**, 1407-1414 (1994).

10. P. Bouthemy, "A Maximum-Likelihood Framework for Determining Moving Edges", *IEEE Trans.*, **PAMI-11**, 499-511 (1989).

11. L. Bedini, I. Gerace, E. Salerno, A. Tonazzini, "Models and Algorithms for Edge-Preserving Image Reconstruction", in *Advances in Imaging and Electron Physics*, **97**, 85-189, P.W. Hawkes, Ed., Academic (1996).