Identifying multiple motions from optical flow *

Alessandra Rognone, Marco Campani, and Alessandro Verri

Dipartimento di Fisica dell'Università di Genova
Via Dodecaneso 33, 16146 Genova, Italy

Abstract. This paper describes a method which uses optical flow, that is, the apparent motion of the image brightness pattern in time-varying images, in order to detect and identify multiple motions. Homogeneous regions are found by analysing local linear approximations of optical flow over patches of the image plane, which determine a list of the possibly viewed motions, and, finally, by applying a technique of stochastic relaxation. The presented experiments on real images show that the method is usually able to identify regions which correspond to the different moving objects, is also rather insensitive to noise, and can tolerate large errors in the estimation of optical flow.

1 Introduction

Vision is a primary source of information for the understanding of complex scenarios in which different objects may be moving non-rigidly and independently. Computer vision systems should be capable of detecting and identifying the image regions which correspond to single moving objects and interpreting the viewed motions in order to interact profitably with the environment. This capability could also be usefully employed to drive the focus of attention and track moving objects in cluttered scenes.

The relative motion of the viewed surfaces with respect to the viewing camera produces spatial and temporal changes in the image brightness pattern which provide a vast amount of information for segmenting the image into the different moving parts [1,2]. As the image motion of nearby points in space which belong to the same surface are very similar, optical flow, i.e., the apparent motion of the image brightness pattern on the image plane [3], is a convenient representation of this information. In addition, simple interpretations of first order spatial properties of optical flow make possible meaningful qualitative and quantitative descriptions of the relative viewed motion which are probably sufficient for a number of applications [2,4-7]. This paper proposes a method, which is based on optical flow, for the detection and identification of multiple motions from time-varying images.

The proposed method consists of three steps. In the first step, a number of linear vector fields which approximate optical flow over non-overlapping squared patches of the image plane are computed. In the second step, these linear vector fields are used to produce a list of the "possible" viewed motions, or labels. Finally, in the third step, a label, that is, a possible motion, is attached to each patch by means of a technique of stochastic relaxation. The labeling of image patches depending on the apparent motion

* This work has been partially funded by the ESPRIT project VOILA, the Progetto Finalizzato Robotica, the Progetto Finalizzato Trasporti (PROMETHEUS), and by the Agenzia Spaziale Italiana. M.C. has been partially supported by the Consorzio Genova Ricerche. Clive Prestt kindly checked the English.

by means of relaxation techniques was first proposed in [2,8]. The presented method has several very good features. Firstly, although accurate pointwise estimates of optical flow are difficult to obtain, the spatial coherence of optical flow appears to be particularly well suited for a qualitative characterisation of regions which correspond to the same moving surface independently of the complexity of the scene. Secondly, even rather cluttered scenes are segmented into a small number of parts. Thirdly, the computational load is almost independent of the data. Lastly, the choice of the method for the computation of optical flow is hardly critical since the proposed algorithm is insensitive to noise and tolerates large differences in the flow estimates.

The paper is organised as follows. Section 2 discusses the approximation of optical flow in terms of linear vector fields. In Section 3, the proposed method is described in detail. Section 4 presents the experimental results which have been obtained on sequences of real images. The main differences between the proposed method and previous schemes are briefly discussed in Section 5. Finally, the conclusions are summarised in Section 6.

2 Spatial properties of optical flow

The interpretation of optical flow over small regions of the image plane is often ambiguous [9]. Let us discuss this fact in some detail by looking at a simple example of a sequence of real images.

Fig. 1A shows a frame of a sequence in which the camera is moving toward a picture posted on the wall. The angle between the optical axis and the direction orthogonal to the wall is 30°. The optical flow which is obtained by applying the method described in [10] to the image sequence and relative to the frame of Fig. 1A is shown in Fig. 1B. It is evident that the qualitative structure of the estimated optical flow is correct. It can be shown [7] that the accuracy with which the optical flow of Fig. 1B and its first order properties can be estimated is sufficient to recover quantitative information, like depth and slant of the viewed planar surface. The critical assumption that makes it possible to extract reliable quantitative information from optical flow is that the relative motion is known to be rigid and translational.

In the absence of similar "a priori" information (or in the presence of more complex scenes) the interpretation of optical flow estimates is more difficult. In this case, a local analysis of the spatial properties of optical flow could be deceiving. Fig. 1C, for example, shows the vector field which has been obtained by dividing the image plane in 64 non-overlapping squared patches of 32×32 pixels and computing the linear rotating vector field which best approximates the optical flow of Fig. 1B over each patch. Due to the presence of noise and to the simple spatial structure of optical flow, the correlation coefficient of this "bizarre" local approximation is very high. On a simple local (and deterministic) basis there is little evidence that the vector field of Fig. 1B is locally expanding. However, a more coherent interpretation can be found by looking at the distributions of Fig. 1D. The squares locate the foci of expansion of the linear expanding vector fields which best approximate the estimated optical flow in each patch, while the crosses locate the centers of rotation of the rotating vector field which have been used to produce the vector field of Fig. 1C. It is evident that while the foci of expansion tend to clusterise in the neighbourhood of the origin of the image plane (identified by the smaller frame), the centers of rotation are spread around. This observation lies at the basis of the method for the identification of multiple motion which is described in the next Section.

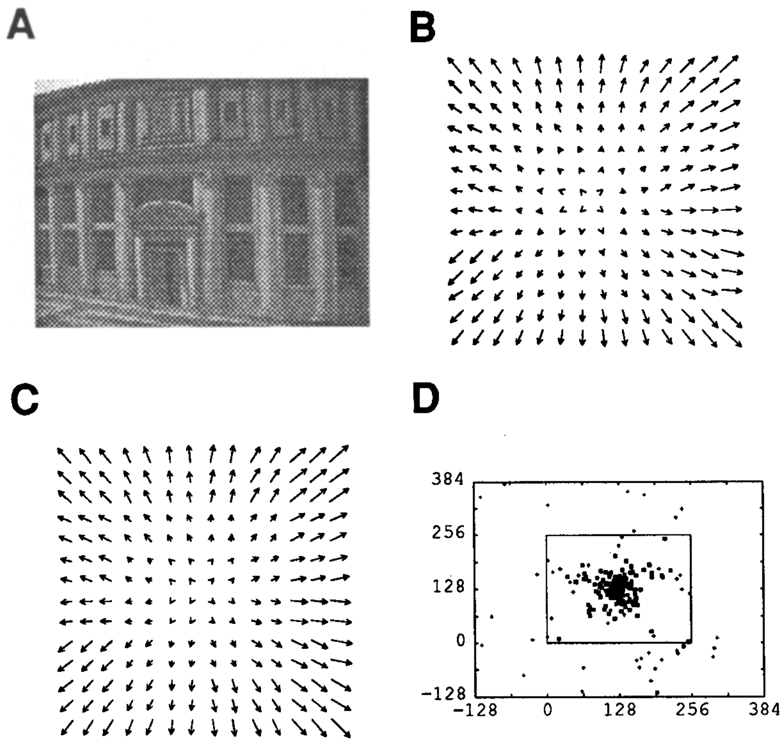

Fig. 1. A) A frame of a sequence in which the viewing camera is moving toward a picture posted on the wall. The angle between the normal vector to the wall and the optical axis is 30°. B) The optical flow computed by means of the method described in [10] associated with the frame of A). C) The optical flow which is obtained by dividing the field of view in 64 squared patches (32×32 pixels each) and computing the linear rotating field which best approximates the optical flow of B) in each patch (in the least mean square sense). D) Distributions of the foci of expansion (squares) and centers of rotation (crosses) of the linear expanding and clockwise rotating vector fields respectively which lie within an area four times larger than the field of view (identified by the solid frame).

3 A method for detecting multiple motions

In this Section a method for detecting and identifying multiple motions from optical flow is proposed. The three main steps of the method are discussed separately by looking at an example of a synthetic image sequence.

3.1 Computing linear approximations of optical flow

Fig. 2A shows a frame of a computer generated sequence in which the larger sphere is translating toward the image plane, while the smaller sphere is translating outward the image plane and the background is rotating. The optical flow relative to the frame of Fig. 2A, and computed through a procedure described in [10], is shown in Fig. 2B. In order to identify the different motions in Fig. 2B, the first step of the method analyses the first order spatial properties of optical flow. The optical flow is divided into patches of fixed size and the expanding (EVF), contracting (CVF), clockwise (CRVF) and anticlockwise (ARVF) rotating, and constant (TVF) vector fields which best approximate the optical flow in each patch s_i, $i = 1, ..., N$, are computed. Roughly speaking, this is equivalent to reducing the possible 3D motions to translation in space with a fairly strong component along the optical axis (EVF and CVF), rotation around an axis nearly orthogonal to the image plane (CRVF and ARVF), and translation nearly parallel to the image plane (TVF). This choice, which is somewhat arbitrary and incomplete, does not allow an accurate recovery of 3D motion and structure (the shear terms, for example, are not taken into account), but usually appears to be sufficient in order to obtain a qualitative segmentation of the viewed image in the different moving objects (see Section 4).

As a result of the first step, five vectors $\mathbf{x}_{s_i}^j$, $j = 1, ..., 5$, are associated with each patch s_i: the vector $\mathbf{x}_{s_i}^1$, position over the image plane of the focus of expansion of the EFV; $\mathbf{x}_{s_i}^2$, position of the focus of contraction of the CVF; $\mathbf{x}_{s_i}^3$, position of the center of the CRVF; $\mathbf{x}_{s_i}^4$, position of the center of the ARVF, and the unit vector $\mathbf{x}_{s_i}^5$, parallel to the direction of the TVF.

3.2 Determining the possible motions

In order to produce a list of the "possible" motions in the second step, global properties of the obtained EVFs, CVFs, CRVFs, ARVFs, and CVFs are analysed. This step is extremely crucial, since the pointwise agreement between each of the computed local vector fields and the optical flow of each patch usually makes it difficult, if not impossible, to select the most appropriate label (see Section 2). Figs. 2C and D respectively show the distribution of the foci of expansion and contraction, and centers of clockwise and anticlockwise rotation, associated with the EVFs, CVFs, CRVFs, and ARVFs of the optical flow of Fig. 2B. A simple clustering algorithm has been able to find two clusters in the distribution of Fig. 2C, and these clusters clearly correspond to the expansion and contraction along the optical axis of Fig. 2B. The same algorithm, applied to the distribution of the centers of rotation (shown in Fig. 2D), reveals the presence of a single cluster in the vicinity of the image plane center corresponding to the anticlockwise rotation in Fig. 2B. On the other hand, in the case of translation, the distribution of the unit vectors parallel to the directions of the TVFs is considered (see Fig. 2E). For the optical flow of Fig. 2B the distribution of Fig. 2E is nearly flat indicating the absence of preferred translational directions. Therefore, as a result of this second step, a label l is attached to each "possible" motion which can be characterised by a certain cluster of points $\mathbf{x}_{s_i}^{c(l)}$, where $c(l)$ equals $1, ..., 4$, or 5 depending on l. In the specific example of Fig. 2, one label of expansion, one of contraction, and one of anticlockwise rotation, are found.

3.3 Labeling through deterministic relaxation

In the third and final step, each patch of the image plane is assigned one of the possible labels by means of an iterative relaxation procedure [11]. The key idea is that of defining a suitable energy function which not only depends on the optical flow patches but also on the possible motions, and reaches its minimum when the correct labels are attached to the flow patches. In the current implementation, the energy function is a sum extended over each pair of neighbouring patches in which the generic term $u(s_i, s_j)$, where s_i and s_j are a pair of neighbouring patches, is given by the formula

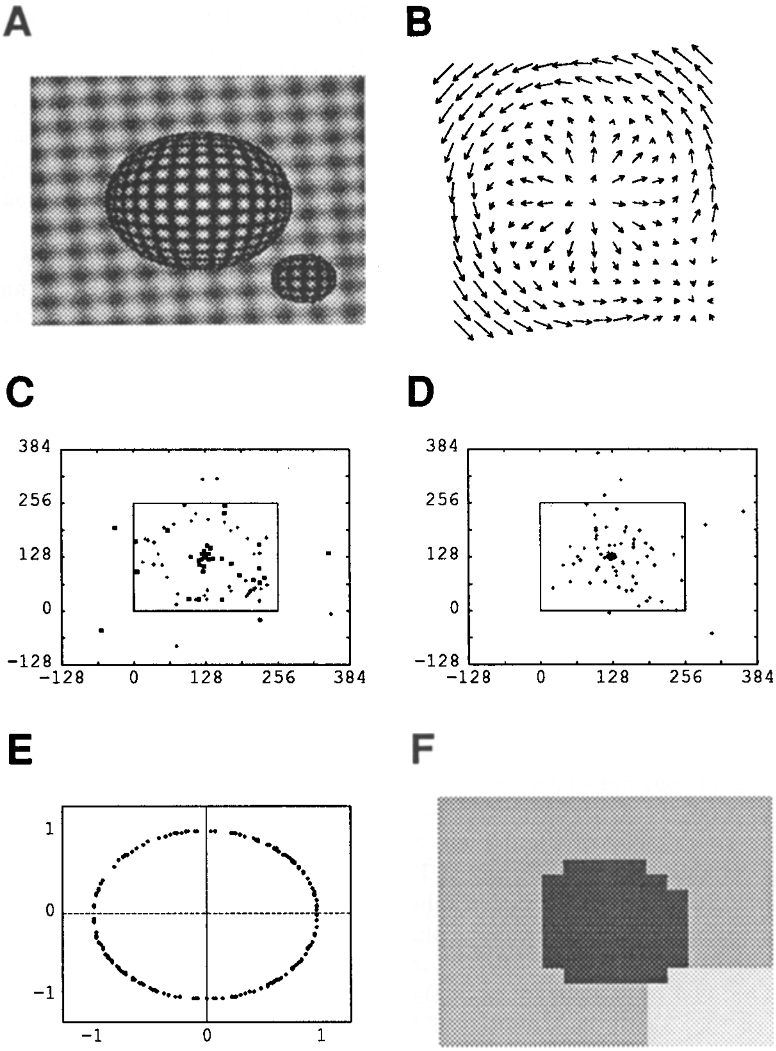

Fig. 2. A) A frame of a synthetic sequence in which the larger sphere is translating toward the image plane, while the smaller sphere is moving away and the background is rotating anticlockwise. B) The corresponding optical flow computed by means of the method described in [10]. C) Distributions of the foci of expansion (squares) and contraction (crosses) of the EVFs and CVFs respectively which lie within an area four times larger than the field of view (identified by the solid frame). D) Distribution of the centers of anticlockwise rotation of the ARVFs. E) Distribution of the directions of the TVFs on the unit circle. F) Colour coded segmentation of the optical flow of B) obtained through the algorithm described in the text.

$$u(s_i, s_j) = \left(\|\mathbf{x}_l - \mathbf{x}_{s_i}^{c(l)}\| + \|\mathbf{x}_l - \mathbf{x}_{s_j}^{c(l)}\|\right) \delta_{l_i = l_j} \tag{1}$$

where $\mathbf{x}_l$ is the center of mass of the cluster corresponding to the label l, and $\delta = 1$ if the labels of the two patches, l_i and l_j respectively, equal l, otherwise $\delta = 0$. The relaxation procedure has been implemented through an iterative deterministic algorithm in which, at each iteration, each patch is visited and assigned the label which minimises the current value of the energy function, keeping all the other labels fixed. The procedure applied to the optical flow of Fig. 2B, starting from a random configuration, produces the colour coded segmentation shown in Fig. 2F after twenty iterations. From Fig. 2F, it is evident that the method is able to detect and correctly identify the multiple motions of the optical flow of Fig. 2B. Extensive experimentation indicates that the deterministic version usually converges on the desired solution. This is probably due to the fact that, for the purpose of detecting multiple motions, the true solution can be approximated equally well by nearly optimal solutions.

To conclude, it has to be said that the profile of the segmented regions can be suitably modeled by adding *ad hoc* terms to the energy (or "penalty functions") which tend to penalise regions of certain shapes. The choice of the appropriate penalty functions reflects the available "a priori" knowledge, if any, on the expected shapes. In the current implementation, in which no "a priori" knowledge is available, only narrow regions have been inhibited (configurations in which in a square region of 3×3 patches there are no five patches with the same label are given infinite energy).

4 Experimental results on real images

Let us now discuss two experiments on real images. Fig. 3A shows a frame of a sequence in which the viewing camera is translating toward the scene while the box is moving toward the camera. The optical flow associated with the frame of Fig. 3A is shown in Fig. 3B. From Fig. 3B it is evident that the problem of finding different moving objects from the reconstructed optical flow is difficult. Due to the large errors in the estimation of optical flow, simple deterministic (and local) procedures which detect flow edges, or sharp changes in optical flow, are doomed to failure. In addition, the viewed motion consists of two independent expansions and even in the presence of precisely computed optical flow, no clear flow edge can be found as the flow direction in the vicinity of the top, right side, and bottom of the box agrees with the flow direction of the background. Fig. 3C shows the distribution of the foci of expansion associated with the EVFs computed as described above. Two clusters are found which correspond to the (independent) motion of the camera and of the box of Fig. 3A. On the contrary, no clusters are found in the other distributions. Therefore, it can be concluded that, at most, two different motions (mainly along the optical axis) are present in the viewed scene. The colour coded segmentation which is obtained by applying the third step of the proposed method is shown in Fig. 3D. It is evident that the algorithm detects and correctly identifies the two different motions of the viewed scene.

In the second experiment (Fig. 4A), a puppet is moving away from the camera, while the plant in the lower part of Fig. 4A is moving toward the image plane. The optical flow associated with the frame of Fig. 4A is reproduced in Fig. 4B. As can be easily seen from Fig. 4C both the distributions of the foci of expansion (squares) and contraction (crosses) clusterise in the neighbourhood of the origin. No cluster has been found in the other distributions, which is consistent with the optical flow of Fig. 4B. The segmentation which is obtained by applying the relaxation step is shown in Fig. 4D.

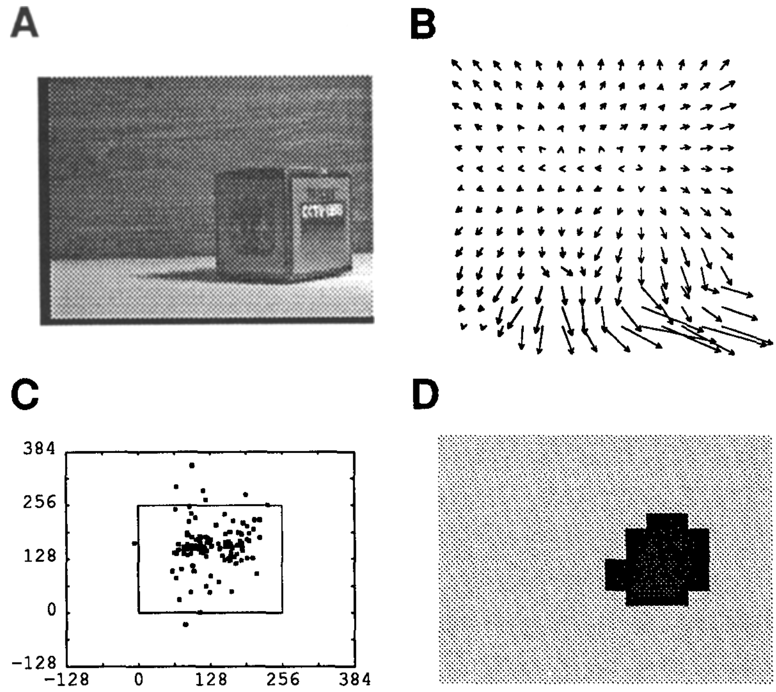

Fig. 3. A) A frame of a sequence in which the box is translating toward the camera, while the camera is translating toward an otherwise static environment. B) The corresponding optical flow computed by means of the method described in [10]. C) Distribution of the foci of expansion of the EVFs. D) Colour coded segmentation of the optical flow of B) obtained through the algorithm described in the text.

This example clarifies the need for two distinct labels for expansion and contraction (and, similarly, for clockwise and anticlockwise rotation). The energy term of Eq. 1, which simply measures distances between singular points, would not be sufficient to distinguish between expanding and contracting patches. In order to minimise the number of parameters which enter the energy function, it is better to consider a larger number of different local motions than to add extra-terms to the right-hand-side of Eq. 1.

To summarise, the proposed method appears to be able to detect multiple motion and correctly segment the viewed image in the different moving objects even if the estimates of optical flow are rather noisy and imprecise.

5 Differences from previous methods

It is evident that the presented method is very different from the deterministic schemes which attempt to identify multiple motions by extracting flow edges [12-13]. Important similarities, instead, can be found with the technique proposed in [2]. Firstly, the same mathematical machinery (stochastic relaxation) is used. Secondly, in both cases first

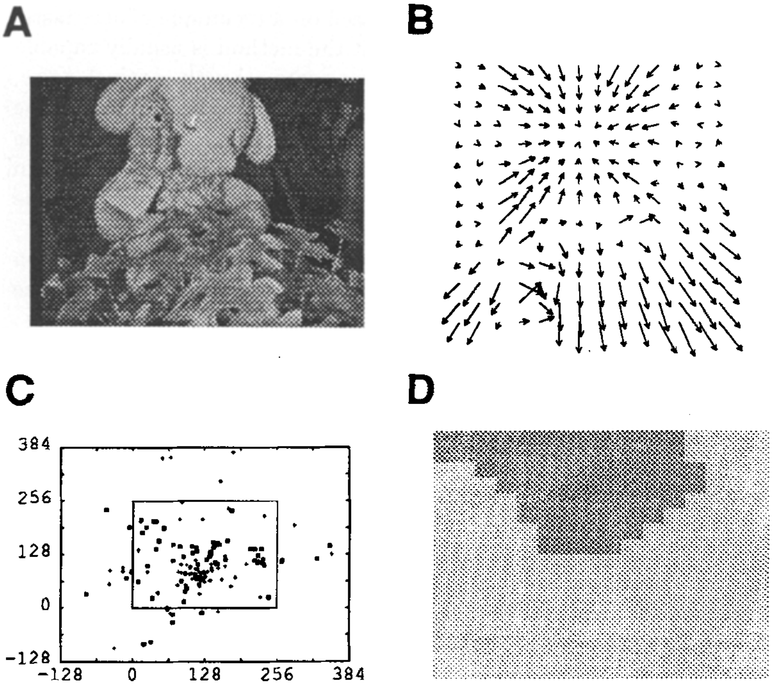

Fig. 4. A) A frame of a sequence in which the puppet is moving away from the camera, while the plant is translating toward the image plane. B) The corresponding optical flow computed by means of the method described in [10]. C) Distribution of the foci of expansion (squares) and contraction (crosses) of the EVFs and CVFS respectively. D) Colour coded segmentation of the optical flow of B) obtained through the algorithm described in the text.

order spatial properties of optical flow, such as expansion and rotation, are employed to determine the different types of motion. However, the two methods are basically different. In [2] regions are segmented and only at a later stage local spatial properties of optical flow are used to interpret the viewed motions. The possible motions are data-independent and the resolution is necessarily fairly low. On the contrary, the method described in the previous Section computes the possible motions first and then identifies the regions which correspond to the different moving objects. Consequently, the number of labels remains small and stochastic relaxation always runs efficiently. In addition, since the possible motion are data-dependent, the resolution is sufficiently high to allow for the detection of "expansion within expansion" (see Fig. 3D) or the determination of arbitrary direction of translation.

6 Conclusion

In this paper a method for the detection and identification of multiple motions from optical flow has been presented. The method, which makes use of linear approximations of optical flow over relatively large patches, is essentially based on a technique of stochastic relaxation. Experimentation on real images indicates that the method is usually capable of segmenting the viewed image into the different moving parts robustly against noise, and independently of large errors in the optical flow estimates. Therefore, the technique employed in the reconstruction of optical flow does not appear to be critical. Due to the coarse resolution at which the segmentation step is performed, the proposed algorithm only takes a few seconds on a Sun SPARCStation for a 256×256 image, apart from the computation of optical flow.

To conclude, future work will focus on the extraction of quantitative information on the segmented regions and will be biased to the theoretical (and empirical) study of the local motions which must be added in order to increase the capability of the method.

References

1. Adiv, G. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE Trans. Pattern Anal. Machine Intell.* **7** (1985), 384-401.
2. Francois, E. and P. Bouthemy. Derivation of qualitative information in motion analysis. *Image and Vision Computing* **8** (1990), 279-288.
3. Gibson, J.J. *The perception of the Visual World.* (Boston, Houghton Mifflin, 1950).
4. Koenderink, J.J. and Van Doorn, A.J. How an ambulant observer can construct a model of the environment from the geometrical structure of the visual inflow. In *Kibernetic 1977*, G. Hauske and E. Butendant (Eds.), (Oldenbourg, Munchen, 1977).
5. Verri, A., Girosi, F., and Torre, V. Mathematical properties of the two-dimensional motion field: from Singular Points to Motion Parameters. *J. Optical Soc. Amer. A* **6** (1989), 698-712.
6. Subbarao, M. Bounds on time-to-collision and rotational component from first order derivatives of image flow. *CVGIP* **50** (1990), 329-341.
7. Campani, M., and Verri, A. Motion analysis from first order properties of optical flow. *CVGIP: Image Understanding* in press (1992).
8. Bouthemy, P. and Santillana Rivero, J. A hierarchical likelihood approach for region segmentation according to motion-based criteria. In *Proc. 1st Intern. Conf. Comput. Vision* London (UK) (1987), 463-467.
9. Adiv, G. Inherent ambiguities in recovering 3D motion and structure from a noisy flow field. *Pattern Anal. Machine Intell.* **11** (1989), 477-489.
10. Campani, M. and A. Verri. Computing optical flow from an overconstrained system of linear algebraic equations. *Proc. 3rd Intern. Conf. Comput. Vision* Osaka (Japan) (1990), 22-26.
11. Geman, D., Geman, S., Graffigne, C., and P. Dong. Boundary detection by constrained optimization. *IEEE Trans. Pattern Anal. Machine Intell.* **12** (1990), 609-628.
12. Thompson, W.B. and Ting Chuen Pong. Detecting moving objects. *IJCV* **4** (1990), 39-58.
13. Verri, A., Girosi, F., and Torre, V. Differential techniques for optical flow. *J. Optical Soc. Amer. A* **7** (1990), 912-922.