Computation of Coherent Optical Flow by Using Multiple Constraints

Massimo Tistarelli
University of Genoa
Department of Communication, Computer and Systems Science (DIST)
Laboratory for Integrated Advanced Robotics (LIRA - Lab)
Via Opera Pia 13 - 16145 Genoa, Italy
E-mail: *tista@dist.unige.it*

Abstract

The optical flow constitutes one of the most widely adopted representations to define and characterize the evolution of image features over time. In order to compute the velocity field, it is necessary to define a set of constraints on the temporal change of image features. In this paper we consider the implications in using multiple constraints arising from multiple data points. The first step is the analysis of differential constraints and how they can be applied, locally, to compute the image velocity. This analysis allows to relate each constraint to a particular gray level pattern.
This approach is extended to multiple image points, allowing also the characterization of the temporal behaviour of the image features and to detect erroneous measurements due to occlusions, depth discontinuities or shadows.

Several experiments are presented from real image sequences.

1 Introduction

The computation of the optical flow field has been studied since many years, and using different approaches. There are many examples in the literature, where optical flow is computed by means of several constraint equations applied to many image points. Minimization and least squares are the most widely applied mathematical tools to solve this kind of problems.
Not always the constraints are obtained by applying the same equation to multiple points, like in [1, 2, 3, 4, 5, 6], but also by defining multiple constraints for each image point, either based on a set of differential equations [7, 8, 9, 10, 11], or obtained by applying the same set of equations to different functions which are related to the local image brightness [12, 13, 14]. It is worth noting that many researchers have explicitly addressed the well known problem of occlusions, in the computation of smooth flow fields. Ben-Tzvi et al. [1] and also Bartolini et al. [2] defined algorithms to avoid the problems of classical least squares techniques, by finding the correct intersections of constraint lines in the velocity space with a method based either on the combinatorial Hough transform, or on a multi windowing technique. Variational techniques have been applied by Nesi [4]. Schnorr [15] proposed a method to deal with flow discontinuities, based on the decomposition of the image plane into disjoint sets. The minimization of the quadratic functional relative to the optical flow field, is performed on the basis of the hypothesized qualitative structure of the flow field within each set.
Otte and Nagel [8] discuss a method, based on the least squares solution of an over-constrained system of linear equations, which relates flow discontinuities with the output of the χ^2 test.

In a preceding paper, the flow constraint equations have been analyzed in terms of the response of the differential operators to different intensity patterns. This analysis allowed to understand which is the best combination of constraints at a given image point. The aim of this paper is to extend this approach to the temporal domain, by considering multiple image points. This extension allows, not only to compute the optical flow, but also to verify the spatial coherence of the measurements, and consequently detect wrong hypotheses on the original constraints, like depth discontinuities or shadows.

2 Motion and optical flow

It is now well known that the aperture problem is a "false problem" [16, 17]. In fact, it can be easily overcome as soon as enough "structure" is present in the image brightness. Assuming the flow field to be locally constant, the time derivative of the gradient can be used, yielding two equations in the two velocity components:

$$\frac{d}{dt}\nabla E = \vec{0} \tag{1}$$

where $E(x, y, t)$ is the image brightness of the point (x, y) at time t. Originally, Tetriack and Pastor [9] and also recently Tistarelli and Sandini [10], considered a locally constant flow model (not implying any particular movement of the objects in space), adding the brightness constancy equation [18] to obtain three

0-8186-7042-8/95 $4.00

equations in two unknowns:

$$\frac{d}{dt}E = 0 \qquad \frac{d}{dt}\nabla E = \vec{0} \qquad (2)$$

with the assumption that the values of the spatial derivatives of the optical flow are equal to zero for all the image points. It is worth noting that this corresponds to a first order approximation of the optical flow field with a locally constant vector field.

3 Using multiple constraints

Considering equation (2) it can be rewritten explicitly:

$$\begin{bmatrix} E_x & E_y \\ E_{xx} & E_{xy} \\ E_{yx} & E_{yy} \end{bmatrix} \cdot \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} E_t \\ E_{xt} \\ E_{yt} \end{bmatrix} \qquad (3)$$

Instead of taking all three equations at the same time, it is possible to consider three equation pairs separately and study their stability. This is done simply looking at the matrices:

$$\mathbf{M}_1 = \begin{bmatrix} E_x & E_y \\ E_{xx} & E_{xy} \end{bmatrix} \; \mathbf{M}_2 = \begin{bmatrix} E_{xx} & E_{xy} \\ E_{yx} & E_{yy} \end{bmatrix} \; \mathbf{M}_3 = \begin{bmatrix} E_x & E_y \\ E_{yx} & E_{yy} \end{bmatrix} \quad (4)$$

This concept can be extended considering any set of functions of the image brightness [14, 13, 12].

Every constraint defines a line in the velocity space (u, v), and the solution is defined by the intersection among the constraint lines. Each intersection, in the velocity space, is characterized by a particular moving gray level distribution. Therefore, the expected error applying a particular equation pair, is directly related to the local structure of the image brightness [19].

It is now evident that not all the equations are "well tuned" to "sense" the temporal variation of every intensity pattern. But, rather, each equation is best suited to compute the motion of a particular intensity distribution [19].

In this approach, the most probable solution is obtained by discarding the intersection corresponding to the lowest determinant and computing the "center of mass" of the remaining intersection points, where the "mass" of each point is the value of the respective matrix determinant. The velocity is computed as:

$$\vec{V} = \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} N_h^u + N_j^u \\ N_h^v + N_j^v \end{bmatrix} \cdot \frac{1}{D_h + D_j} \qquad (5)$$

where $N_i^u = b_i^1 M_i^{22} - b_i^2 M_i^{12}$ and $N_i^v = b_i^2 M_i^{11} - b_i^1 M_i^{21}$ are the numerators of the expression for $\vec{V}_i = (u_i, v_i)$, and $D_i = \det \mathbf{M}_i$.

4 Integration of data from multiple points

The application of multiple constraints can be extended to multiple data points, by assuming the flow field to be locally constant. This can be performed either by applying a least squares estimator to the resulting set of equations, or by using the same methodology applied to combine the constraints relative to each single pixel. The correct velocity vector is determined, in the velocity space, as the center of mass of the cloud of intersections, relative to each pixel within the considered neighborhood.

As a matter of facts, rarely the flow field will be constant, but this assumption is well approximated if the considered neighborhood is small compared to the size of the low-pass operator applied to blur the images. Because the blurring is performed both in space and time, the final effect is to minimize the differences between neighbouring pixels, thus enforcing the local constancy assumption.

What does happen if the flow field is not constant (or smooth)?

The flow field is not constant whenever:

- the relative velocity between the observer and the scene has a rotational component;
- the relative velocity between the observer and the scene is only translational, but includes a component along the direction of the optical axis.
- the relative velocity between the observer and the scene has only a translational component with direction parallel to the image plane, but the imaged scene is not constituted by one single planar surface parallel to the image plane.

Even though the requirements to obtain a constant velocity field are difficult to meet, it is generally possible to locally approximate the optical flow with a constant vector field. In general, abrupt changes in the velocity field are due to depth discontinuities, occlusions and shadows. If the considered pixel, or region, corresponds to a portion of surface in space under one of these conditions, then the flow field can not be computed locally, neither by considering few neighbouring pixel, nor by applying only the constraints relative to one single pixel [15, 4].

The advantage of considering multiple pixels stems from the possibility to verify if the constancy of the flow field is a good approximation of the real flow field. In fact, if this is the case, the flow constraints relative to neighbouring points should define approximately the same velocity vector.

It is possible to evaluate if the flow field is locally constant (or smooth), by analyzing the distribution of the intersection points of the constraint lines relative to all the pixels within the considered neighborhood. Each pixel will give rise to a set of constraints and, consequently, few intersection points will be selected in the velocity space. All the intersection points will form a cloud of points in the velocity space. If the velocity field is locally constant or smooth, a very small cloud is expected, because all the pixels will have the same velocity. On the other hand, if the cloud is large, or the intersection points are spread apart, then the pixels exhibit different velocities.

In order to analyze the distribution of the intersection points, the modulus of the mean $E(\vec{V})$ and of the

variance $\sigma(\vec{V})$ of the distribution have been defined:

$$E(\vec{V}) = \left| \frac{1}{N} \sum_{i=0}^{N-1} \vec{d}_i \right| \quad \vec{d}_i = \begin{bmatrix} |u_i - u| \\ |v_i - v| \end{bmatrix} \quad (6)$$

$$\sigma(\vec{V}) = \left| \frac{1}{N} \sum_{i=0}^{N-1} \left[\vec{d}_i - \vec{\delta} \right]^2 \right| \quad \vec{\delta} = \frac{1}{N} \sum_{i=0}^{N-1} \vec{d}_i \quad (7)$$

where N is the number of selected intersection points, $\vec{V} = (u, v)$ is the velocity vector computed from the set of intersection points, and $\vec{V}_i = (u_i, v_i)$ is the velocity vector defined by each intersection point in the velocity space. The vector $\vec{\delta}$ represents the mean distance of the set of intersection points from the top of the resulting velocity vector $\vec{V}$.

The value of $E(\vec{V})$ defines how sparse are the intersection points in the velocity space, but the variance $\sigma(\vec{V})$ defines the regularity of the distribution. The mean, seems to be the most important to determine whether there is a drastic change within the gray level structure. The variance does not convey the same information, but it is not clear yet, to which extent it could be used to analyze the smoothness in the variation of the image brightness over time.

By setting a threshold on the value of $E(\vec{V})$ it is possible to determine if the velocity field is locally constant or, conversely, if the current image location corresponds to an occlusion or depth discontinuity or a moving shadow.

4.1 Including linear variations of velocity

By considering multiple pixels it is also possible to include a first order, non constant, term in the flow constraints (2). Linear variations of the optical flow can be expressed as [3]:

$$\vec{V}(\vec{\mathbf{x}}) = \vec{V}(\vec{\mathbf{x}}_0) + \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} \cdot \begin{bmatrix} \Delta_x \\ \Delta_y \end{bmatrix} \quad (8)$$

By substituting this expression into (2) one obtains:

$$\begin{aligned} E_x\,\alpha &+ E_y\,\beta = -E_t \\ E_{xx}\,\alpha &+ E_{yx}\,\beta + E_x\,u_x + E_y\,v_x = -E_{tx} \\ E_{xy}\,\alpha &+ E_{yy}\,\beta + E_x\,u_y + E_y\,v_y = -E_{ty} \end{aligned} \quad (9)$$

$$\begin{aligned} \alpha &= (u(\vec{\mathbf{x}}_0) + u_x\,\Delta_x + u_y\,\Delta_y) \\ \beta &= (v(\vec{\mathbf{x}}_0) + v_x\,\Delta_x + v_y\,\Delta_y) \end{aligned}$$

where Δ_x and Δ_y are the displacements between the reference point $\vec{\mathbf{x}}_0 = (x_0, y_0)$, where the velocity vector is computed, and the point $\vec{\mathbf{x}} = (x, y)$, where the derivatives of the image intensity are computed.

In this case, the constraint lines in the velocity space are parametrized by the derivatives of the optical flow. Therefore, every two points give one velocity estimate, where equation (9) are applied to one point and equations (2) are applied to the second point. This is possible because all the measurements are referred to the same velocity vector, at the same point on the image plane, while the derivatives of the image velocity are supposed to be locally constant.

Considering a 3×3 neighborhood, 8 intersections can be computed by taking the constraints from the central pixel and, in turn, from every pixel out of the 8 pixels around. The drawback of this approach is the increased complexity of the equations to be solved for each intersection. In fact, instead of a 2×2 matrix of coefficients $\mathbf{M}_i$, we will obtain a 6×6 matrix for every set of pixels:

$$\begin{bmatrix} E_x^0 & 0 & 0 & E_y^0 & 0 & 0 \\ E_{xx}^0 & E_x^0 & 0 & E_{yx}^0 & E_y^0 & 0 \\ E_{xy}^0 & 0 & E_x^0 & E_{yy}^0 & 0 & E_y^0 \\ E_x^j & E_x^j\,\Delta_x & E_x^j\,\Delta_y & E_y^j & E_y^j\,\Delta_x & E_y^j\,\Delta_y \\ E_{xx}^j & E_{xx}^j\,\Delta_x + E_x^j & E_{xx}^j\,\Delta_y & E_{yx}^j & E_{yx}^j\,\Delta_x + E_y^j & E_{yx}^j\,\Delta_y \\ E_{xy}^j & E_{xy}^j\,\Delta_x & E_{xy}^j\,\Delta_y + E_x^j & E_{yy}^j & E_{yy}^j\,\Delta_x & E_{yy}^j\,\Delta_y + E_y^j \end{bmatrix}$$

$$\cdot \begin{bmatrix} u^0 \\ u_x \\ u_y \\ v^0 \\ v_x \\ v_y \end{bmatrix} = - \begin{bmatrix} E_t^0 \\ E_{tx}^0 \\ E_{ty}^0 \\ E_t^j \\ E_{tx}^j \\ E_{ty}^j \end{bmatrix}$$

where $(u^0, v^0) = \vec{V}(\vec{\mathbf{x}}_0)$, the derivatives of the image brightness E_{hk}^0 are computed at the point $\vec{\mathbf{x}}_0$ and the derivatives E_{hk}^j are computed at the point $\vec{\mathbf{x}}_j$.

The optical flow and its first order derivatives can be computed if the parameter matrix is not singular. This is verified if the intensity gradients at the image points $\vec{\mathbf{x}}_0$ and $\vec{\mathbf{x}}_j$ are not null and not aligned.

5 Experimental results

In figure 1 to 3 several experiments, performed by computing the optical flow using multiple points, are shown. The parameters which have been set in the algorithm are the following:

- σ_s and σ_t are the values of the standard deviation of the Gaussian kernel in space and time;
- Δ is the maximum allowed difference in the values of the highest determinants;
- τ is the threshold on the values of the denominator computing the flow vectors;
- σ_f is the value of the standard deviation of the Gaussian kernel used to smooth the X and Y component of the computed velocity field.

In order to improve visibility the flow vectors have been sub-sampled by taking one vector every 5, along the X and Y image coordinates.

The first experiment, reported in figure 1, has been performed on the "laboratory" sequence. The sequence has been acquired from an active head equipped with two stereo cameras, mounted on a wheeled platform and moving inside our laboratory, while fixating a point on the table in the foreground. On top are the computed raw optical flow (left) and the optical flow smoothed with a Gaussian kernel with

$\sigma_f = 1.2$ (right). The optical flow has been obtained by thresholding the vectors with a value of the mean $E(\vec{V})$ exceeding 0.3 pixels.
The parameters used for the computation are: $\sigma_s = 2.5$, $\sigma_t = 1$, $\Delta = 5\%$, $\tau = 1$.
In the middle, one original image is shown. On the right, an enlargement of the computed velocity vector, in the velocity space, at two different positions (marked with a cross on the original image), are shown. All the intersection points used to computed the vectors are also shown as small dots. On the bottom, are the computed mean (left) and variance (right) for all the image points. The light gray indicates a zero value, all other values range from 1 (dark) to 255 (white). As it can be noticed, most of the higher values of the mean and variance are at the image points corresponding to depth discontinuities.

In figure 2 an experiment performed on the "moving" sequence is shown. The sequence has been acquired inside a room, from a camera mounted on a mobile vehicle moving along a direction slightly inclined with respect to the camera optical axis. The object on the right, in the foreground, is moving along a direction almost orthogonal to the trajectory of the vehicle. On top of figure 2 one original image (left) and the computed raw optical flow (right) are shown. The optical flow has been obtained by thresholding the vectors with a value of the mean $E(\vec{V})$ exceeding 0.2 pixels.
The parameters used for the velocity computation are: $\sigma_s = 2$, $\sigma_t = 1$, $\Delta = 5\%$, $\tau = 0.9$.
In the middle, the optical flow smoothed with a Gaussian kernel with standard deviation σ_f equal to 1 is shown. On the right, an enlargement of one computed velocity vector in the velocity space, corresponding to the position marked with a cross on the original image, is shown. All the intersection points used to compute the vector are also shown as small dots. On the bottom, are the computed mean (left) and variance (right) for all the image points. The light gray indicates a zero value, all other values range from 1 (dark) to 255 (white). It is worth noting that the area corresponding to the moving object in the foreground, has very high values both for the mean and for the variance of the distribution of the intersection points in velocity space. High values in the mean are also reported along the depth discontinuities.

The last experiment has been performed on an image sequence acquired from a camera mounted on a robot arm at the University of Karlsruhe. The camera was moving with a pure 3D translation toward the scene. The objects are static, with the exception of one light marble block, which is translating toward the left edge of the picture. The sequence is composed of 30 images with a resolution of 512x512 pixels and with 8 bits of resolution in intensity. The images have been reduced to 256x256 pixels by cropping a 256x256 window around the moving marble block.
On top of figure 3 the computed raw optical flow (left) and the optical flow smoothed with a Gaussian kernel with standard deviation σ_f equal to 1 (right) are shown. The optical flow has been obtained by thresholding the vectors with a value of the mean $E(\vec{V})$ exceeding 0.5 pixels.
The parameters used for the computation are: $\sigma_s = 2$, $\sigma_t = 1$, $\Delta = 5\%$, $\tau = 1$. In the middle, one original image is shown. On the right, an enlargement of the computed velocity vector, in the velocity space, at two different positions (marked with a cross on the original image), are shown. All the intersection points used to computed the vectors are also shown as small dots. On the bottom, are the computed mean (left) and variance (right) for all the image points. The light gray indicates a zero value, all other values range from 1 (dark) to 255 (white). As it can be noticed, most of the higher values of the mean and variance are at the image points corresponding to depth discontinuities or to occluding edges of the two blocks in the image. High values in the mean are also reported, on the lower left corner, within an area with uniform intensity.

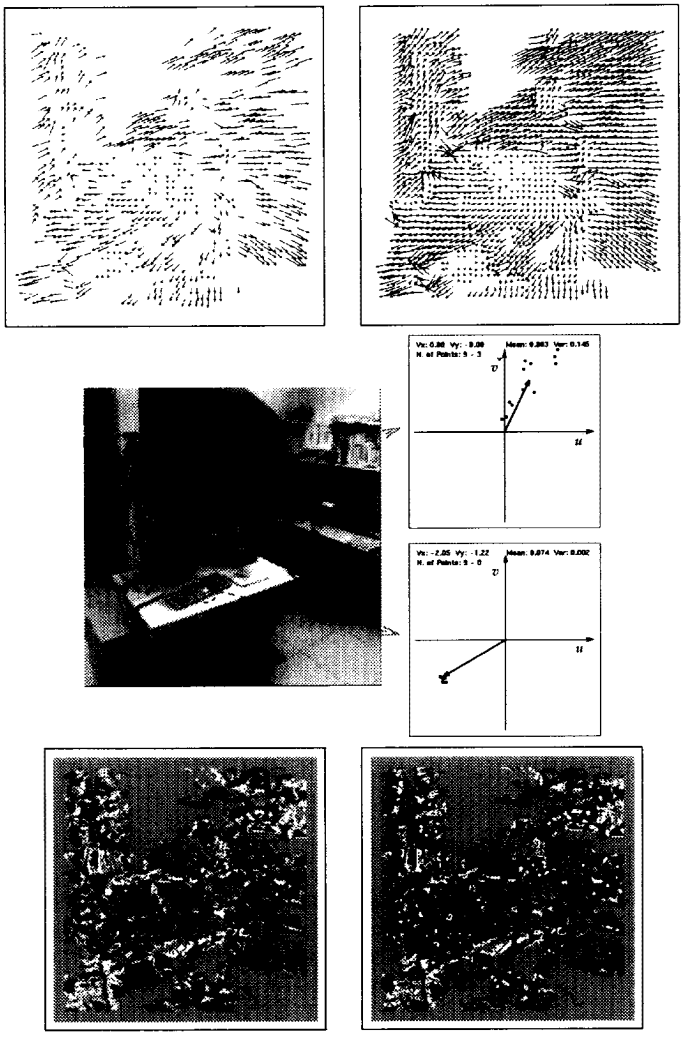

Figure 1: Computation of the optical flow by applying constraints from multiple points on the "laboratory" image sequence.

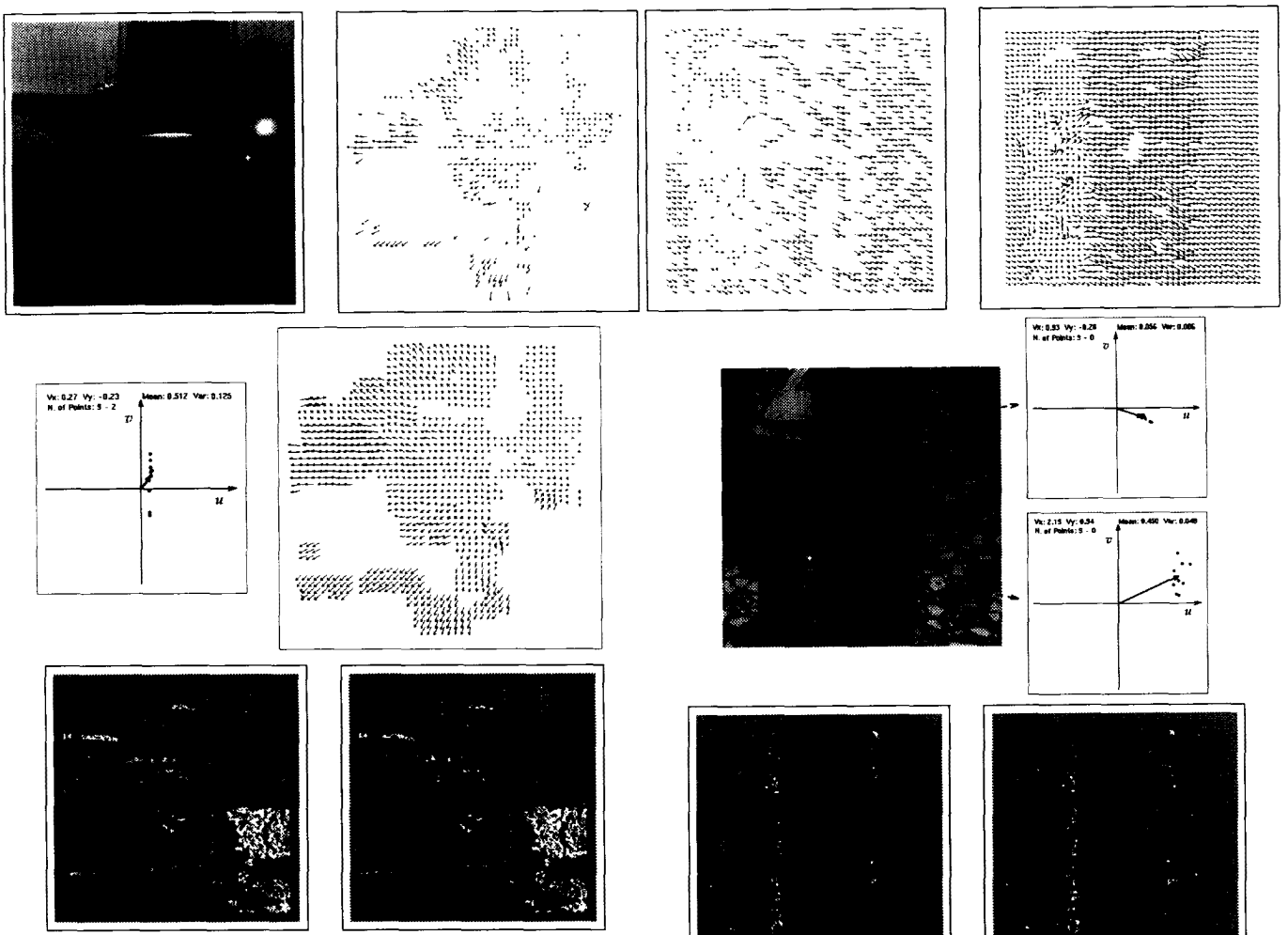

Figure 2: Computation of the optical flow by applying constraints from multiple points on the image sequence containing a moving object.

Figure 3: Optical flow computed by applying constraints from multiple points on the "blocks" image sequence. The sequence has been obtained by extracting a 256x256 window from the original 512x512 images.

6 Conclusion

In this paper we have addressed the problem of combining multiple constraints to compute the optical flow field from an image sequence.

One of the main aspects which has been outlined in this paper is that the response of a given constraint, strictly depends on the local distribution of the image intensity. Therefore, the choice of the constraints to be applied should depend on the local structure of the image brightness and not only on the confidence associated to the measurement. In fact, there are examples where the local image structure does not allow to apply a given constraint at all, or the information obtained is completely wrong. These observations lead to the conclusion that, in order to compute the optical flow field from an image stream, the constraints to be applied to the image sequence should not be chosen only on the basis of the motion to be detected, but also on the local image structure. In fact, it has been shown in the past, that an abundance of equations can be written, depending on the motion constraints we are considering.

Not all the equations are equally suitable for computing the flow field at all image points. We have demonstrated, both analytically and with experiments, that the same equations applied to different brightness structures can give exact or wrong estimates.

The complex nature of the real world, many times makes the assumptions for the velocity computation to fail. This is due to many phenomena, like occlusions, shadows, depth discontinuities or even an excessive velocity of the objects. These conditions can be detected by analyzing the behaviour of the image intensity distribution along space and time. It has been shown that this is obtained by analyzing the distribution of the intersection points obtained from the constraint lines, in the velocity space, relative to neighbouring pixels.

Two simple measurements, namely the mean and the

standard deviation of the distribution of the intersections among the constraint lines, are devised to detect and possibly threshold the flow vectors corresponding to image points where some of the assumptions may be violated.
Many experiments performed on real image sequences demonstrated the validity of the approach.
As a matter of facts, other measurements could be used to characterize the distribution of the constraints in the velocity space. By applying different measurements other properties could be detected, and eventually determine which assumption has been violated. In other words it could be possible to understand if the considered pixel corresponds to a shadow or an occlusion or a depth discontinuity and so on. This analysis could be usefully applied to better understand the nature of the scene for successive vision tasks.

Acknowledgements

This work has been partially funded by the Esprit project VAP, by grants of Italian National Research Council and by a grant of the SMART European network of excellence.
I wish to thank Dr. E. Grosso and Prof. G. Sandini for the stimulating discussions and helpful insights in the development of this work.

References

[1] D. Ben-Tzvi, A. Del Bimbo, and P. Nesi, "Optical flow from constraint lines parametrization", *Pattern Recognition*, vol. 26, no. 10, pp. 1549–61, 1993.

[2] F. Bartolini, V. Cappellini, C. Colombo, and A. Mecocci, "Multiwindow least-squares approach to the estimation of optical flow with discontinuities", *Optical Engineering*, vol. 32, no. 6, pp. 1250–6, 1993.

[3] M. Campani and A. Verri, "Computating optical flow from an overconstrained system of linear algebraic equations", in *Proc. of 3rd IEEE Intl. Conference on Computer Vision*, Osaka, Japan, 1990, pp. 22–6, IEEE.

[4] P. Nesi, "Variational approach to optical flow estimation managing discontinuities", *Image and Vision Computing*, vol. 8, no. 4, pp. 419–39, 1993.

[5] B. G. Schunck, "Image flow segmentation and estimation by constraint line clustering", *IEEE Trans. on PAMI*, vol. PAMI-11, no. 10, pp. 1010–27, 1989.

[6] K. Wohn, L. S. Davis, and P. Thirft, "Motion estimation based on multiple local constraints and non-linear smoothing", *Pattern Recognition*, vol. 16, pp. 563–70, 1983.

[7] M. Otte and H. H. Nagel, "Optical flow estimation: Advances and comparisons", in *Proc. of 3rd European Conference on Computer Vision*, J. O. Eklundh, Ed., Stockholm, Sweden, May 2-6, 1994, pp. 51–60, Springer Verlag.

[8] H. H. Nagel, G. Socher, H. Kollnig, and M. Otte, "Motion boundary detection in image sequences by local stochastic tests", in *Proc. of 3rd European Conference on Computer Vision*, J. O. Eklundh, Ed., Stockholm, Sweden, May 2-6, 1994, pp. 305–316, Springer Verlag.

[9] O. Tretiak and L. Pastor, "Velocity estimation from image sequences with second order differential operators", in *Proc. of 7th IEEE Intl. Conference on Pattern Recognition*. 1984, pp. 16–9, IEEE.

[10] M. Tistarelli and G. Sandini, "Dynamic aspects in active vision", *CVGIP, Special issue on Purposive and Qualitative Active Vision*, Y. Aloimonos, Ed., vol. 56, no. 1, pp. 108–129, July 1992.

[11] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques", *Int. J. of Computer Vision*, vol. also Tech. Rep. RPL-TR-9107, 1993.

[12] V. Markandey and B. E. Flinchbaugh, "Multispectral constraints for optical flow computation", in *Proc. of 3rd IEEE Intl. Conference on Computer Vision*, Osaka, Japan, 1990, pp. 38–41, IEEE.

[13] J. Lai, J. Gauch, and J. Crisman, "Using color to compute optical flow", *Proceedings of the SPIE*, vol. 2056, pp. 186–94, 1993.

[14] R. J. Woodham, "Multiple light source optical flow", in *Proc. of 3rd IEEE Intl. Conference on Computer Vision*, Osaka, Japan, 1990, pp. 42–6, IEEE.

[15] C. Schnorr, "Computation of discontinuous optical flow by domain decomposition and shape optimization", *International Journal of Computer Vision*, vol. 8, no. 2, pp. 153–65, 1992.

[16] S. Uras, F. Girosi, A. Verri, and V. Torre, "A computational approach to motion perception", *Biological Cybernetics*, vol. 60, pp. 79–87, 1988.

[17] H. H. Nagel, "On the estimation of optical flow: Relations between differenet approaches and some new results", *Artificial Intelligence*, vol. 33, pp. 299–324, 1987.

[18] B. K. P. Horn and B. G. Schunck, "Determining optical flow", *Artificial Intelligence*, vol. 17, no. 1-3, pp. 185–204, 1981.

[19] M. Tistarelli, "Multiple constraints for optical flow", in *Proc. of 3rd European Conference on Computer Vision*, J. O. Eklundh, Ed., Stockholm, Sweden, May 2-6, 1994, pp. 61–70, Springer Verlag.