Motion Segmentation from Optical Flow

Alessandro Verri[†], Sergio Uras[‡] & Enrico De Micheli[‡]

[†]Scuola Internazionale Superiore di Studi Avanzati,
Strada Costiera 11, 34014 Grignano (TS), Italy.

[‡] Dipartimento di Fisica dell' Università di Genova,
Via Dodecaneso 33, 16146 Genova, Italy.

This note describes two different methods for motion segmentation from optical flow. In the first method, the Jacobian matrix of the first spatial derivatives of the components of optical flow is used to compute the amount of uniform expansion, pure rotation, and shear at every flow point. In the second description, local properties of optical flow which are invariant for non-singular linear transformation are computed from the trace and the determinant of the matrix itself. Both the methods allow to distinguish between different kinds of motion, like translation, rotation, and relative motion in sequences of time-varying images. Preliminary results show that they can also be useful to identify the different moving objects in the viewed scene.

Time-varying images provide useful information for the understanding of several visual problems. This information, which can be thought of as encoded in optical flow [1] — a dense planar vector field which gives the velocity of points over the image plane —, appears to be essential for important visual tasks like passive navigation and dynamic scene understanding (see [2-5] for example).

In this note, the optical flow computed through a technique which has recently been proposed in ref. [6,7] is used for motion segmentation. Two different local descriptions of image motion that are invariant for orthogonal transformations of coordinates on the image plane (*i.e.*, arbitrary rotation of the viewing camera around the optical axis) are discussed. In the first description, which is obtained by looking at the changing image brightness as a $2D$ deformable body, the Jacobian matrix of the first spatial derivatives of the components of optical flow is used to compute the amount of uniform expansion, pure rotation, and shear at every flow point. The second description focuses on the local properties of optical flow which are invariant for non-singular linear transformation and thus can be inferred from the trace and the determinant of the matrix itself. According to these descriptions, image motion can be segmented in regions where either the average percentage of uniform expansion, pure rotation, or shear is larger than a fixed value, or where the qualitative nature of the eigenvalues do not change.

Experiments on real images show that, in many cases, these regions roughly correspond to the image of the observed moving objects and make it possible to distinguish between different kinds of $3D$ motions. The computation of the percentage of uniform expansion, pure rotation, and shear seems to be less sensitive to noise than the study of the qualitative nature of the eigenvalues of the Jacobian matrix. In the case of translation, which is discussed analytically, the percentage of uniform expansion is usually much larger than the percentages of pure rotation and shear. At the same time, the two eigenvalues of the Jacobian matrix are real and often almost equal. At boundary points, instead, the shear component is larger and the eigenvalues may have opposite sign. Motion segmentation for rotation and relative motion is also discussed. Finally, it is shown that the integration between the presented motion segmentation and other visual cues like intensity edges allows to obtain more accurate image segmentation.

Some conclusions can be drawn from the presented analysis. Firstly, it has been shown that it is possible to obtain motion segmentation from optical flow. Two different techniques based on the study of the Jacobian matrix of optical flow have been implemented which can be used to segment the image plane in regions that allow to distinguish between different kinds of motion, like translation, rotation, and relative motion, and to identify the different moving objects. The presented results complement recent results [8] that have been obtained on qualitative and quantitative properties of the Jacobian matrix at the singular points of optical flow, that is, the points where the flow vanishes. Here, the segmentation and the analysis of the spatial structure of optical flow in the neighborhood of singular points, which were essential for the understanding of the observed $3D$ motion in ref. [8], are obtained easily and reliably from local analysis. In fact, the obtained motion segmentation is useful even when no singular point is found in optical flow. Finally, it appears that the technique which has been used for the computation of optical flow [6,7] is not only adequate for $3D$ motion recovery from singular points of optical flow [6,7,9], but also for motion and object segmentation.

UNIFORM EXPANSION, PURE ROTATION, AND SHEAR

Optical flow can be extracted from image sequences by making suitable assumptions on the time-varying im-

age brightness. Recently [6,7] it has been proposed to compute the optical flow $\vec{u} = (u_1, u_2)$ at the location $\vec{x} = (x_1, x_2)$ on the image plane at the time t by solving the system of algebraic equations for (u_1, u_2)

$$\frac{d\vec{\nabla}E}{dt} = 0 \qquad (1)$$

where $d/dt = \partial/\partial t + \vec{u} \cdot \vec{\nabla}$, $\partial/\partial t$ is the partial temporal derivative, $\vec{\nabla}$ is the $2D$ spatial gradient, and $E = E(x_1, x_2, t)$ is the time-varying image brightness on the image plane. The optical flow which can be obtained from eq. 1 is usually adequate for the understanding of motion and the estimate of motion parameters from singular points [6,7]. However, in order to reliably extract motion information from the singular points of optical flow, or through some other method, it is useful to study the spatial structure of optical flow at other locations.

We now show that the spatial structure of optical flow at every point can be described in terms of three quantities, α, β, and γ, which give the amount of uniform expansion,

pure rotation, and shear in the flow respectively at that point. At every point $\vec{x}$ in a sufficiently small neighborhood of a point $\vec{x}_0$ of the image plane, the optical flow $\vec{u}(\vec{x})$ can be written as

$$\vec{u}(\vec{x}) = \vec{u}(\vec{x}_0) + \mathbf{M}(\vec{x} - \vec{x}_0) \qquad (2)$$

where

$$\mathbf{M}_{ij} = \frac{\partial u_i}{\partial x_j}$$

are the entries of the Jacobian matrix $\mathbf{M}$ of the partial spatial derivatives of the components of $\vec{u}$ computed at $\vec{x}_0$. A theorem of Helmoltz [10] shows that the most general infinitesimal motion of a sufficiently small element of a deformable body (in the $2D$ case) can be decomposed in the sum of a rigid translation, a uniform expansion, a rotation, and two orthogonal component of shear. If the changing image brightness is viewed as the $2D$ deformable body and optical flow as the infinitesimal motion, as suggested in ref. [11,9], it is clear that in eq. 2 the term of rigid translation is given by $\vec{u}(\vec{x}_0)$. The other four terms can be obtained [9] by writing the matrix M
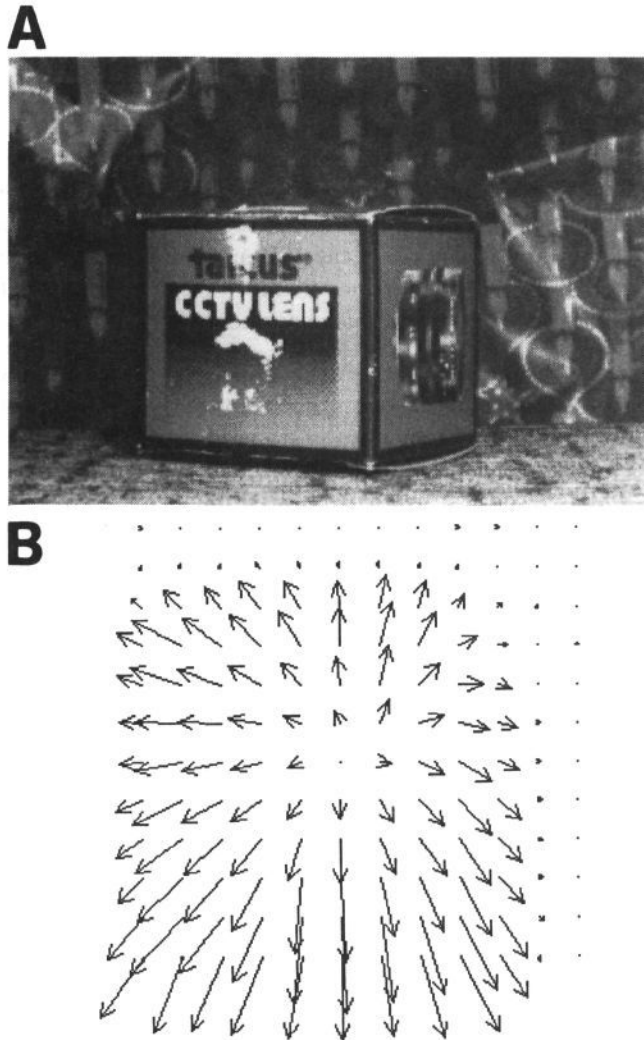
A

B

Figure 1. *Time varying image sequences: case of translation. A) One frame (256 × 256 pixels) of a sequence in which a box is translating toward the viewing camera; B) optical flow computed at the frame in A) according to the procedure described in ref. [7] and subsampled over a 32 × 32 grid for better visualization.*
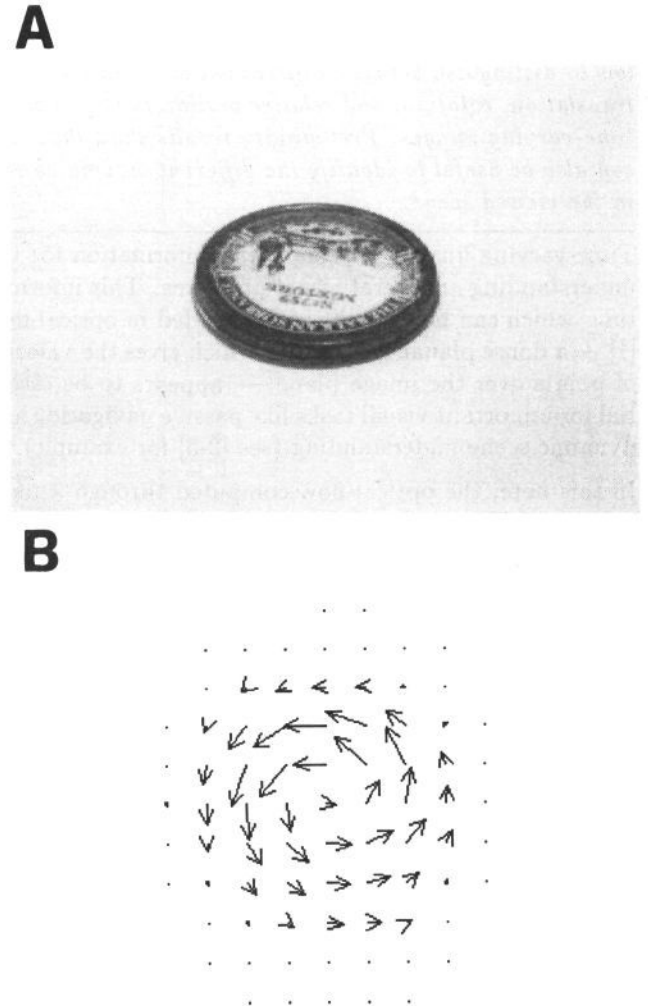
A

B

Figure 2. *Time varying image sequences: case of rotation. A) one frame of a sequence in which a box is rotating on a plane tilted with respect to the image plane; B) optical flow computed at the frame in A) (see legend of Figure 1B).*

as

$$M = \alpha I_1 + \beta I_2 + \gamma_1 I_3 + \gamma_2 I_4$$

where $\alpha = (M_{11} + M_{22})/2$, $\beta = (M_{12} - M_{21})/2$, $\gamma_1 = (M_{11} - M_{22})/2$, $\gamma_2 = (M_{12} + M_{21})/2$, and

$$I_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \quad I_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix};$$

$$I_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}; \quad I_4 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The matrices I_i, $i = 1, ..., 4$ correspond to the uniform expansion, the pure rotation, and the two orthogonal components of shear respectively. It is easy to see that I_1 and I_2 (which are the identity and antisymmetric tensor respectively) are left unchanged for arbitrary orthogonal transformation. Thus α and β are also invariant. Consequently, the amount of uniform expansion and pure rotation, which are independent of the system of coordinates, are intrinsic properties of optical flow. On the contrary, since the two components of shear are not invariant separately, γ_1 and γ_2 do not describe intrinsic properties of the flow. An invariant measure of the amount of shear in optical flow is given by $\gamma = \sqrt{\gamma_1^2 + \gamma_2^2}$. In what follows, therefore, we will refer to γ as the amount of shear in optical flow.

We now show that α, β, and γ can be used for motion segmentation. Figures 1A and 2A show two images of a translating and a rotating object respectively. The corresponding optical flows computed by solving eq. 1 according to the technique described in ref. [7] are shown in Figures 1B and 2B. Let us consider first the case of translation. The shaded region in Figure 3A, which is superimposed to the translating box of Figure 1A, marks the locations on the image plane where the average percentage of uniform expansion in the optical flow of Figure 1B is much larger (more than 70 %) than the average percentages of pure rotation and shear. It can be seen that the shaded region roughly corresponds to the moving object. Clearly, the region grows if the threshold on the average percentage is lowered.

This result can also be discussed analytically by considering a smooth translating surface. From the equation of perspective projection it is easy to obtain

$$\alpha = -\frac{T_3}{X_3} + \frac{1}{2} \frac{\vec{n} \cdot \vec{u}}{\vec{n} \cdot \vec{x}}$$

$$\beta = \frac{1}{2} \frac{\vec{n} \times \vec{u} \cdot \vec{e}_3}{\vec{n} \cdot \vec{x}}$$

$$\gamma = \frac{u}{2} \frac{|\vec{n} \times \vec{e}_3|}{|\vec{n} \cdot \vec{x}|}$$

where T_3 is the component of the translational velocity in the direction parallel to the optical axis and to the unit vector $\vec{e}_3$, X_3 is the distance of a point p on the moving surface from the image plane, and $\vec{n}$ the unit normal vector to the surface at p. The optical flow $\vec{u}$ is to be thought of as a $3D$ vector of norm u and coordinates $(u_1, u_2, 0)$ computed at $\vec{x} = (x_1, x_2, f)$, the image point of p, where f is the focal length.

A
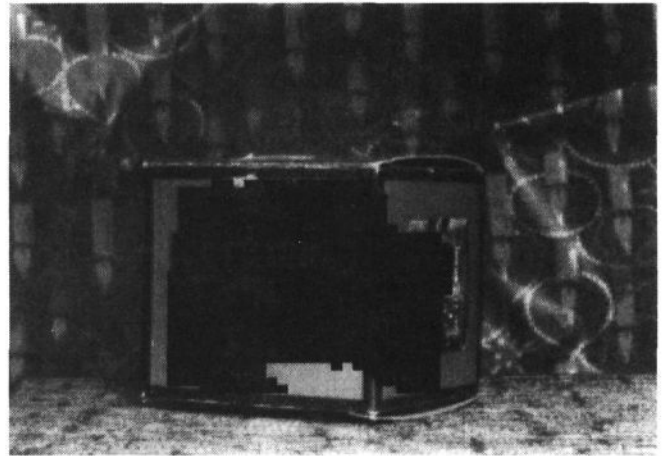
B

Figure 3. *Motion segmentation through the computation of the amount of uniform expansion, pure rotation, and shear. A) the shaded region marks the locations where the average percentage of uniform expansion in the optical flow of Figure 1B is 79%. The percentages of shear and pure rotation were 14% and 7% respectively. The region has been obtained as follows. As a first step, the procedure selects the flow points at which the percentage of uniform expansion, pure rotation, and shear is larger than 30%. Then, the connected regions which contains at least one thousand pixels are picked. Finally, the regions where the average percentage of either one of the three components is larger than 70% are extracted; B) the shaded region marks the locations where the average percentage of pure rotation in the optical flow of Figure 2B is 76%. The percentages of shear and uniform expansion were 21% and 3% respectively. Processing as in A).*

Notice that when the normal to the moving surface is parallel to the image plane, we have $\beta = \gamma = 0$. The component of pure rotation and shear vanishes also at the singular point, that is, where $\vec{u} = 0$, and it is always true that $|\beta| \leq \gamma$. Even if, in principle, neither β nor γ are bounded, in most situations we have that α is much larger than β or γ. This is clearly true when the

angle θ between $\vec{n}$ and $\vec{x}$ is small (θ is the angle between the tangent plane to the moving surface and the image plane), because β and γ vanish with θ^2, whereas the term T_3/X_3 in α does not depend on θ. For large θ (up to 70 degrees), typical velocities and distances ($T_3 \simeq 1cm/frame$ and $X_3 \simeq 100cm$), focal length of about one thousand pixels and average magnitude of optical flow of a few pixels per frame, it can be easily checked that the ratio between α and either one of β and γ is still some units.

The case of rotation is more complex, since even the optical flow of a rotating planar surface can be very different from a pure rotation on the image plane. However, in the simple case shown in Figure 2A, the tangent plane to the rotating box is orthogonal to the rotation axis. Thus, we expect to find a large component of pure rotation in the flow of Figure 2B. The shaded region in Figure 3B marks the location where the average percentage of pure rotation in the optical flow of Figure 2B is above 70 %. The quality of the resulting motion segmentation is comparable with the segmentation obtained in the case of translation.

EIGENVALUES OF THE JACOBIAN MATRIX

Let us describe an alternative method for motion segmentation. When optical flow is seen as a smooth vector field [12,8], useful information about qualitative and quantitative properties of motion can be recovered, by using the theory of dynamical systems, from the Jacobian matrix at the singular points of optical flow, which are the points where the flow vanishes. It has been shown [8] that the spatial structure of optical flow in the neighborhood of a singular point, which is described by the trace and the determinant of the Jacobian matrix at the singular point, make it possible to distinguish between different types of motion and allows to recover relevant motion parameters, like time-to-collision for translation and angular velocity for rotation.

Therefore, in order to study the Jacobian matrix at the singular points, it may be useful to extend the qualitative analysis to the neighboring points and, in fact, to every flow point. The Jacobian matrix is expected to vary continuously over optical flow and thus the qualitative properties of the trace and the determinant of the matrix should remain the same over regions of the image plane.

Let us consider again the optical flow of the translating box in Figure. 1B. The shaded region in Figure 4A marks the locations where the trace, $Tr\mathbf{M}$, and the determinant, $Det\mathbf{M}$, of the Jacobian matrix satisfy (up to some allowed deviation) the relationship
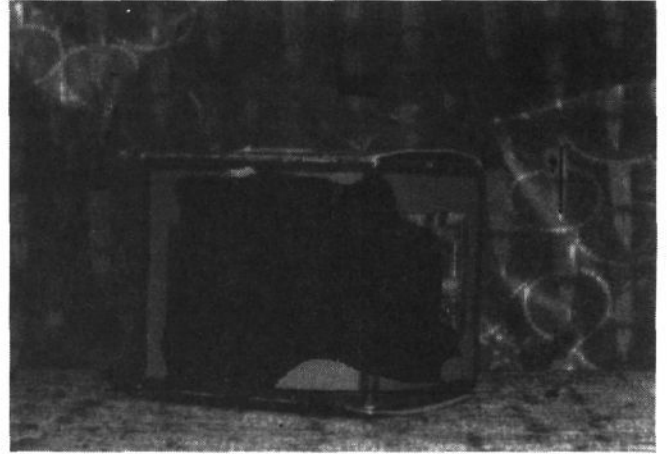
$$Det\mathbf{M} \simeq \frac{Tr^2\mathbf{M}}{4} \qquad (3)$$

Eq. 3 implies that the eigenvalues of the Jacobian matrix are equal. It is straightforward to see that this property is typical of a translating plane which is parallel to the image plane. In fact, for the eigenvalues λ_1 and λ_2 of $\mathbf{M}$ in the general case of a smooth translating surface we have

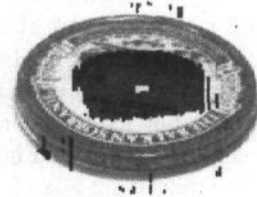$$\lambda_1 = -\frac{T_3}{X_3}$$

$$\lambda_2 = -\frac{T_3}{X_3} + \frac{\vec{n} \cdot \vec{u}}{\vec{n} \cdot \vec{x}}$$

A

B

Figure 4. Motion segmentation through the computation of the trace and the determinant of the Jacobian matrix of optical flow. A) the shaded region marks the locations where the relative difference between DetM and $Tr^2M/4$ is within 25%, that is, where the eigenvalues of M are roughly equal; B) the shaded region marks the locations where DetM is larger than $Tr^2M/4$ and the relative difference is greater than 25%, that is, where the eigenvalues of M are complex.

Thus, λ_1 and λ_2 are always real. Moreover, λ_1 equals λ_2 at the singular point and when the normal vector $\vec{n}$ to the moving surface is orthogonal to the image plane. According to the same argument of the previous section, it can be concluded that $\lambda_1 \simeq \lambda_2$ at most locations and for typical values of motion and structure parameters. Since λ_1 is always positive, the eigenvalues can have opposite signs only if λ_2 can assume negative values. It can be shown [13] that this happens if the moving surface is very far from the viewer and the tangent plane to the surface forms a wide angle with the image plane. In practice, however, this case is of little interest because

it is usually very difficult to correctly compute optical flow at locations where the tangent plane to the moving surface is nearly orthogonal to the image plane.

Concluding, from the qualitative properties of the eigenvalues of the Jacobian matrix and from the average percentage of uniform expansion, it can easily be understood whether the observed motion is a translation independently of the location of the singular point. Figure 4B shows the same analysis for the optical flow of Figure 2B in the case of rotation. In this case we expect to find that the eigenvalues are always complex or that

$$Det\mathbf{M} < \frac{Tr^2\mathbf{M}}{4} \qquad (4)$$

A

B

Figure 5. Motion and image segmentation in a sequence of relative motion. A) One frame representing a relative motion in which an oscilloscope was translating outward and an electrical device was translating toward the viewing camera; B) optical flow computed at the frame in A); (see legend of Figure 1B).

The shaded region in Figure 4B marks the locations where the inequality 4 holds true. In both the cases of Figure 4, it has to be noticed that the segmentation which can be obtained by means of constraints like eq. 3 and inequality 4 is rather sensitive to noise. This fact

implies that the choice of the parameters and the accuracy in the computation of optical flow are more critical for this method than for the method of section 2.

Finally, let us consider a more complex dynamic scene where motion segmentation appears to be almost essential. Figure 5A shows an image of a relative motion between an oscilloscope, which is translating outward, and an electric device, which is translating toward the
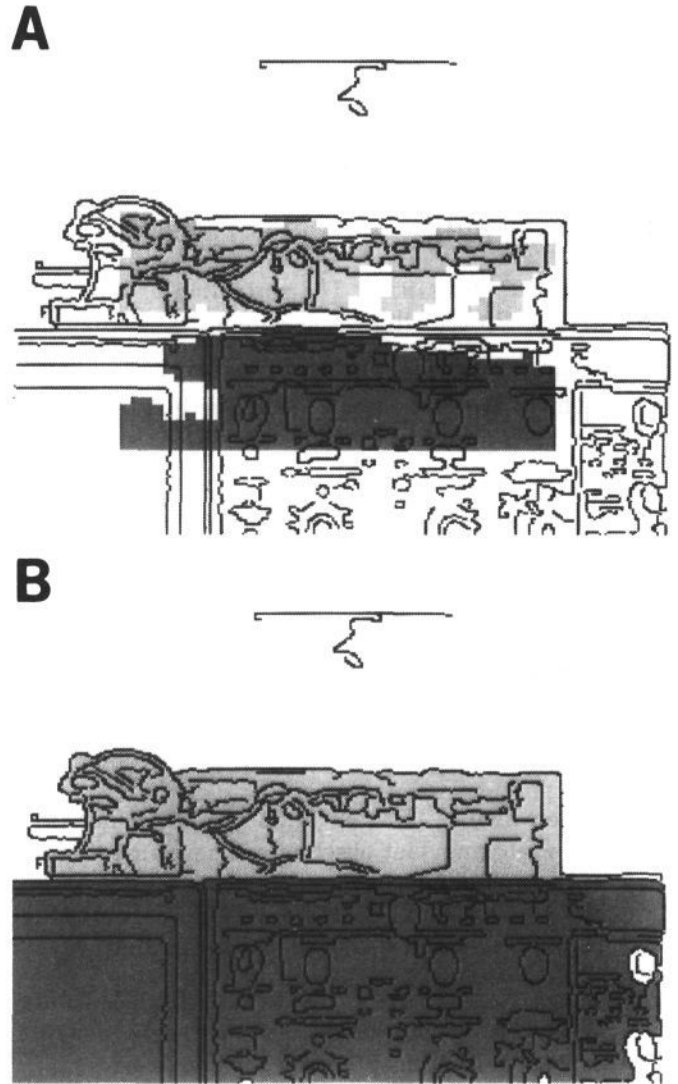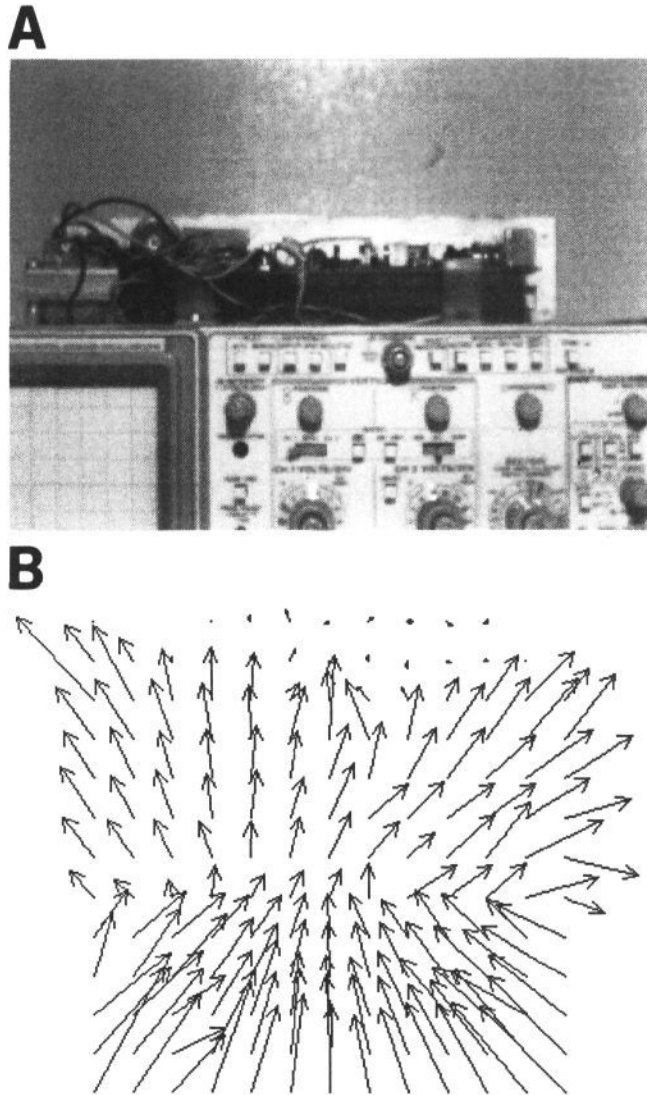
A

B

Figure 6. A) The shaded regions mark the locations where the average percentage of uniform expansion (light grey) and uniform contraction (dark grey) of the optical flow in Figure 5B is larger than 70%. Processing as in legend of Figure 3A. Intensity edges of the image in Figure 5A have been superimposed. Edge points have been extracted by means of gaussian convolution with unit standard deviation (mask size = 5 pixels) and non-maxima suppression of the magnitude of the spatial gradient. Then, edge points are chained by means of a standard contour following procedure; B) a more accurate segmentation can be obtained by means of a rather simple procedure. The regions in A) are recursively enlarged until an edge point is reached. The interior of a closed edge chain which turns out to be surrounded by an enlarged region is annexed to the region itself.

viewing camera. The computed optical flow is shown in Figure 5B. Figure 6A reproduces the regions where the average percentage of uniform expansion (light grey) and uniform contraction (dark grey) is larger than 70 %. Intensity edges of the image in Figure 5A have been superimposed to compare the resulting motion segmentation vs. the correct segmentation. A very accurate segmentation (see Figure 6B) can be obtained by a simple region growing algorithm, described in the legend of Figure 6, which takes into account intensity edges.

REFERENCES

1. **Gibson, J.J.** 1950. *The perception of the Visual World*. Boston: Houghton Mifflin.

2. **Longuet-Higgins, H.C. & K. Prazdny.** 1980. "The interpretation of a moving retinal image". In *Proc. R. Soc. Lond. B* 208: 385-397.

3. **Dreschler, L. & H.-H. Nagel.** 1982. "Volumetric model and 3D trajectory of a moving car derived from monocular TV frame sequences of a street scene". In *Comput. Vision Graph. Image Process.* 20: 199-228.

4. **Ullmann, S.** 1983. "Recent computational studies in the interpretation of structure from motion". In *Human and Machine Vision*, A. Rosenfeld & J. Beck (Eds.), Academic Press, New York.

5. **Koenderink, J.J. & A.J. van Doorn.** 1977. "How an ambulant observer can construct a model of the environment from the geometrical structure of the visual inflow". In G.Hauske and E. Butendant (Eds.) *Kibernetic 1977* (Oldenbourg, Munchen).

6. **Uras, S., Girosi, F., Verri, A., & V. Torre.** 1988. "A computational approach to motion perception". *Biological Cybernetics*, 60: 79-87.

7. **Baraldi, P., De Micheli, E., Radonich, G., & V. Torre.** 1989. "The recovery of motion and depth from optical flow". *Proceedings of the Optical Society Topical Meeting on Image Understanding and Machine Vision*, Cape Cod MA.

8. **Verri, A., Girosi, F. & V. Torre.** 1989. "Mathematical Properties of the 2D Motion Fields: from Singular Points to Motion Parameters". *Proceedings of the IEEE Workshop on Visual Motion*, Irvine CA.

9. **Girosi, F., Verri, A., & V. Torre.** 1989. "Constraints in the Computation of Optical Flow". *Proceedings of the IEEE Workshop on Visual Motion*, Irvine CA.

10. **Helmholtz, A.** 1858. "Über Integrale der hydrodynamischen Gleichungen, welche den Wirbelwegungen entsprechen". *Crelles J.* 55: 25.

11. **Koenderinck, J.J.** 1986. "Optic Flow". *Vis. Res.* 26: 161-168 .

12. **Verri, A. & T. Poggio.** 1989. "Motion Field and Optical Flow: Qualitative Properties". *IEEE Trans. on PAMI*, 11, PAMI-5, 490-498.

13. **Verri, A. & F. Aicardi.** 1989. In preparation.