Motion and Depth from Optical Flow

Patrizia Baraldi[†], Enrico De Micheli[‡] & Sergio Uras[‡]

† Lab. di Bioingegneria, Facolta' di Medicina,
Universita' di Modena, Modena, Italy

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

Passive navigation of mobile robots is one of the challenging goals of machine vision. This note demonstrates the use of optical flow, which encodes the visual information in a sequence of time varying images [1], for the recovery of motion and the understanding of the three dimensional structure of the viewed scene. By using a modified version of an algorithm, which has recently been proposed to compute optical flow, it is possible to obtain dense and accurate estimates of the true 2D motion field. Then these estimates are used to recover the angular velocity of the viewed rigid objects. Finally it is shown that, when the camera translation is known, a coarse depth map of the scene can be extracted from the optical flow of real time varying images.

The navigation of a robot in any environment requires the knowledge of the motion of the robot relative to the environment, the three dimensional structure of the scene and the motion parameters of the object moving in the scene. These informations can be obtained by using active sensors and/or by passive vision, provided by cameras mounted on the robot.

In this note it is shown how passive vision can be used to recover depth when the camera on the robot is translating and angular velocity when the camera looks at rotating objects. The proposed technique first computes the optical from a sequence of time varying images by using a modification of the algorithm recently proposed [2,3]. The angular velocity can be obtained by exploiting mathematical properties of the 2D motion field [5]. Depth is obtained from the computed optical flow, using an equation already proposed by many authors (Horn 1987 , Tommasi personal communication).

THE COMPUTATION OF OPTICAL FLOW

The motion of objects in the viewed scene at every time t defines a $3D$ velocity field, which is transformed by the imaging device of a T.V. camera in a $2D$ vector field $\vec{v} = (\vec{v}_1(x,y), \vec{v}_2(x,y))$, usually called the $2D$ motion field.

It has been shown [3] that a $2D$ vector field close to $\vec{v}$, usually called the optical flow $\vec{u}$, can be obtained by solving the vector equation

$$\frac{d}{dt} \text{ grad } E = 0 \qquad (1)$$

where d/dt indicates the total time derivative (or the Eulerian derivative) and grad E is the spatial gradient of the image brightness $E(x,y,t)$. From eq. 1 the optical flow $\vec{u}$ can be computed as

$$\vec{u} = -\mathbf{H}^{-1} \frac{\partial}{\partial t} \text{ grad } E \qquad (2)$$

where $\mathbf{H}$ is the Hessian matrix of $E(x,y)$.

The relation between the optical flow $\vec{u}$ and the true $2D$ motion field $\vec{v}$ is:

$$\vec{u} = \vec{v} + \mathbf{H}^{-1} \left(\mathbf{J}\vec{v}^{\,T} \text{ grad } E - \text{ grad } \frac{dE}{dt} \right) \qquad (4)$$

where $\mathbf{J}\vec{v}^{\,T}$ indicates the transpose of the Jacobian matrix of $\vec{v}$.

It is evident from eq. 2 that numerical stability of the computation of the optical flow requires a robust inversion of the matrix $\mathbf{H}$, which is garanteed when det $\mathbf{H}$ is large, and the conditioning number $\mathbf{C_H}$ of the matrix $\mathbf{H}$ is close to one [7]. From eq. 4 it is evident that when the term in brackets is bounded the optical flow $\vec{u}$ is close to the true $2D$ motion field $\vec{v}$ when the entries of the matrix $\mathbf{H}^{-1}$ are small. Since $\mathbf{H}$ is symmetric the conditioning number $\mathbf{C_H}$ is equal to $\lambda_{max}/\lambda_{min}$ where λ_{max} and λ_{min} are the largest and smallest of the two real eigenvalues of $\mathbf{H}$. As a consequence when det $\mathbf{H}$ is large and $\mathbf{C_H}$ is close to one, the two eigenvalues λ_{max} and λ_{min} will be both large ensuring that:

i. the computation of the optical flow from eq. 2 is numerically robust;

ii. the optical flow $\vec{u}$ is usually close to the true $2D$ motion field $\vec{v}$, with the exception of those locations in the image where $\mathbf{J}\vec{v}^{\,T}$ or $\frac{dE}{dt}$ are very large.

Therefore the conditions det $\mathbf{H}$ large, and $\mathbf{C_H} \sim 1$ usually garantee good recovery of the $2D$ motion field.

We now describe the procedure used to compute the optical flow from a sequence of time-varying images, such as the one shown in Figures 1A and 3A.
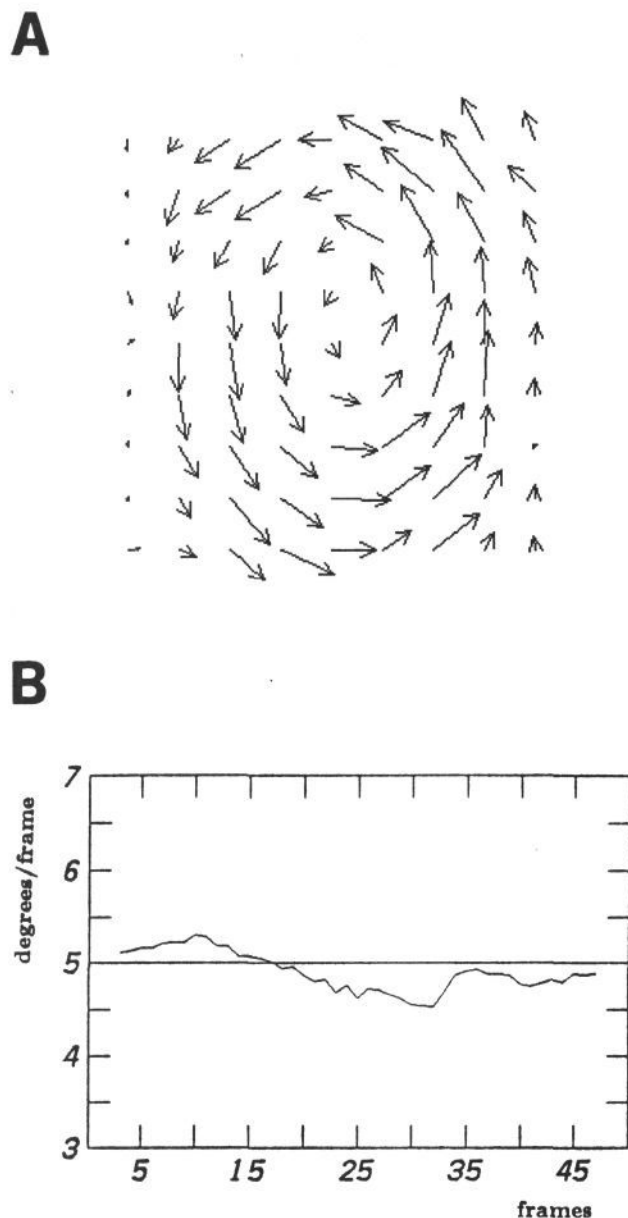
A

B

A

B
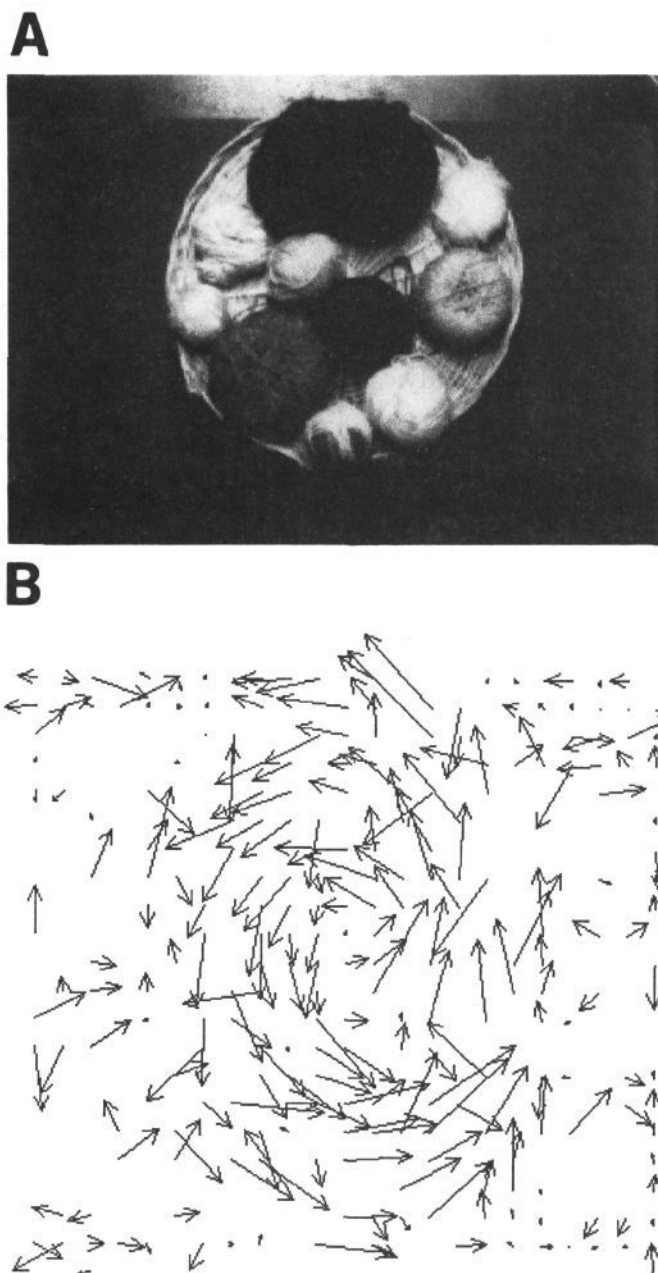
Figure 1. Computation of optical flow for a rotating object. A) One frame of a sequence of an indoor scene. B) Row optical flow obtained with spatial gaussian smoothing with mask size = 5 pixels, temporal gaussian smoothing with mask size = 9 pixels, Det H > 0.05 and C_H < 10.

Figure 2. Computation of angular velocity. A) Optical flow of Figure 1B after filling-in and smoothing with gaussian filter with mask size = 21. B) Angular velocity computed from the optical flow of Figure 2A versus time (in degree/frame).

The optical flow obtained by solving eq. 2 is first computed at each location. Many vectors with an erroneous magnitude or direction are clearly present. By choosing only those vectors, obtained when the computation of H^{-1} is numerically robust (det H large and $C_H \sim 1$) a sparse, but almost exact optical flow is obtained (Figures 1B and 3B).

ANGULAR VELOCITY FROM OPTICAL FLOW

In order to recover a dense and meaningful optical flow it is desirable to fill in the empty areas of the optical flow by using a filling-in procedure. The optical flow is then smoothed by the convolution with a gaussian

symmetrical filter and the results are shown in Figures 2A and 4A. In Figure 1A a scene in which the rotation axis was about parallel with the optical axis of the viewing camera is displayed. In Figure 2A the immobile point of the rotation, i.e. the point perspective projection of a point which lies on the rotation axis, is clearly present and the angular velocity ω is equal to $|\lambda|$, where λ is the eigenvalue of the $J\vec{v}$ matrix computed to the immobile point (if the optical axis is parallel to the rotation axis we have $|\lambda_1| = |\lambda_2|$) [3]. Figure 2B compares the true angular velocity (solid straight line) and the computed angular velocity for the sequence of images of Figure 1A. The agreement between the true and computed angular velocity depends on the texture of the scene. When the
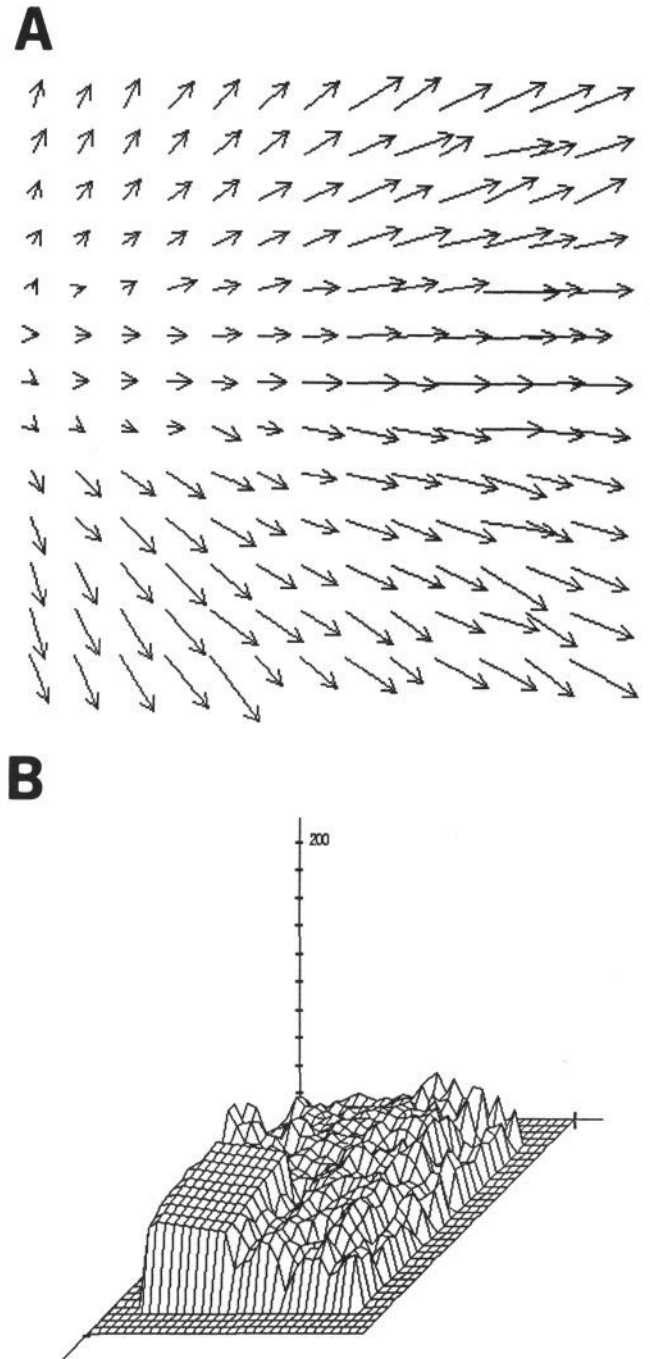
206

A

B

A

B
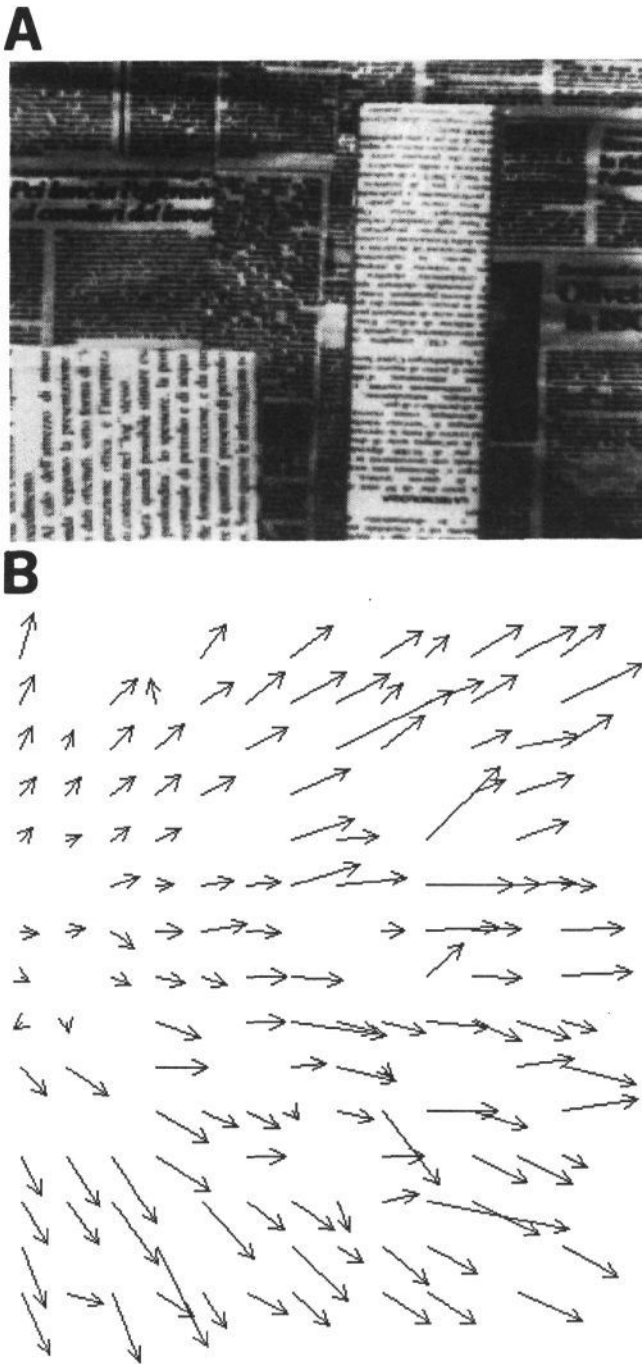
Figure 3. Computation of optical flow for ego-motion. A) One frame of a sequence representing two books stacks acquired by a camera moving towards them. B) Row optical flow obtained with a gaussian smoothing with mask size = 7 pixels, no temporal smoothing, Det H > 0.1 and conditioning number C_H < 5. The focus of expansion is visibly located at the left side of the image plane; the angle between the direction of translation and the optical axis was of about 10 degrees.

Figure 4. Computation of a depth map. A) Optical flow of Figure 3B after filling-in and smoothing with a gaussian filter with mask size = 17. B) Depth map obtained by the optical flow of Figure 2C. The true distances from the image plane were 37 cm, 56 cm and 72 cm for the books stack in the lower left side, for the one on the right side and for the background respectively. Computed mean values for the three regions were 35 ± 5 cm, 61 ± 9 cm and 62 ± 13 cm.

viewed scene is densely textured the accuracy can be as high as 95% (the mean accuracy in the sequence of Figure 1A is 95.4%)

DEPTH FROM OPTICAL FLOW

Optical flow can also be used to recover depth from motion by using the formula

$$Z = v_{\text{camera}} \frac{D}{V} \qquad (5)$$

where v_{camera} is the velocity of the moving camera, D is the distance of the point (x, y) on the image plane from the focus of expansion F_e, V is the amplitude of the flow in (x, y), and Z is the depth of the point in the scene projected in (x, y). Figure 3A shows an image from a sequence taken by a camera translating towards

two book stacks at different depth. The books and the background were covered with newspaper sheets in order to increment the texture of the scene. To avoid the computation of depth near the focus of expansion, where the optical flow values are noisy, the angle between the direction of translation and the optical axis was set to about 10 degrees. Consequently the focus of expansion lied near to the image boundary. Figure 3B reproduces the computed optical flow and Figure 4A the optical flow after the filling-in and smoothing procedures. When depth is averaged over the results obtained from a few frames we obtain the map shown in Figure 4B. The results obtained for the higher book stack of the scene are in good agreement with the true depth, whereas parts of the 3D structure of the lower one are noisy and almost indistinguishable from the background (see legend for numerical details).

This note presents an algorithm which is adequate to compute a dense optical flow from which it is possible to recover motion information and depth. The computation of optical flow occurs in two steps: in the first step by using eq. 2 a row optical flow is obtained out of which only the reliable displacement vectors are selected; in the second step a dense optical flow is obtained by filling in holes of the optical flow produced in the first step. Different filling-in procedures can be used with different results. The proposed algorithm seems very suitable to solve the problem of motion analysis in machine vision.

REFERENCES

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

2. **Uras, S., Girosi, F., Verri, A., & V. Torre.** 1988. "A computational approach to motion perception". *Biological Cybernetics*, 60: 69-87 .

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

4. **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).

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

6. **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.

7. **Lanczos, C.** 1961. *Linear Differential Operators*. London: D. Van Nostrand Company.

8. **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.