MVA '92 IAPR Workshop on Machine Vision Applications Dec. 7-9,1992, Tokyo

ESTIMATION AND INTERPRETATION OF OPTICAL FLOW FIELDS FOR COUNTING MOVING OBJECTS

Alberto Del Bimbo[‡], Paolo Nesi[‡*], Jorge L. C. Sanz[*†]

[‡] Department of Systems and Informatics, University of Florence, Italy.

[*] Computer Science Department, IBM Almaden Research Center, California, USA.

[†] Computer Research and Advanced Applications Group, IBM Argentina.

Abstract

The main goal of sequence analysis is the motion estimation of moving objects which are present in the scene. One of the most important approaches for motion estimation is based on the estimation of an approximated projection of 3D motion on the image plane, which is usually called optical flow. The collection in time of these optical flows can be useful to understand the object structure and behavior. In this paper, the problem of counting people by interpreting optical flow fields in space and time is presented. Experimental results of counting people who are going in/out of a public bus are reported.

1 Introduction

One of the most important approaches for motion estimation is based on the estimation of an approximated projection of 3D motion on the image plane, which is usually called optical flow [1], [2], [3], [4].

The gradient-based approach for optical flow estimation provides solutions to this problem starting from the observation of brightness changes in the image plane [1], [6], [7], [8], [9], [10]. The optical flow, in general, differs from the perspective projection of the 3-D motion on the image plane which is usually called "velocity field" or "motion field" [11], [2], [4]. However, the estimation of an approximate velocity field, such as the optical flow, can be very useful for many applications.

The estimation of motion is a typical problem of short term analysis. On the other hand, the long term analysis can be made collecting the results obtained in time with short term techniques. Typical problems of long term analysis are: the object tracking and predict position, motion understanding, recognition by motion [5].

In this paper the long-term problem of counting people is analyzed. Specifically, experimental results are presented for the problem of determining the number of people who are going in/out of a public bus. In this case, shapes in the image frames, which represent bodies of moving people, cannot be regarded as rigid objects. Moreover, due the position of the inspecting camera, these are typically not fully focused and exceeds the image boundaries. For these reasons, matching-based techniques for motion analysis cannot be profitably used. In order to overcome the above referred problems, optical flow estimation is used in the paper to determine the local velocity field on the image plane. An interpretation of the optical flow fields in time and space is hence performed to count people.

The paper is organized as follows: in Section 2, the fundamentals of the gradient-based approach for optical flow estimation are reported. In Section 3, the solutions adopted for optical flow estimation, are presented. In Section 4, the method for interpreting the optical flow field in order to count the people which pass under the video camera is proposed. Conclusions are drawn in Section 5.

2 Gradient-Based Approach

Most of the motion estimation techniques presented in the literature to evaluate the optical flow use an equation called Optical Flow Constraint (OFC). The definition of the OFC derives from the observation that the changes in the image brightness $E(x(t), y(t), t)$ with respect to t, can be denoted by:

$$\frac{dE}{dt} = \frac{\partial E}{\partial x}\frac{dx}{dt} + \frac{\partial E}{\partial y}\frac{dy}{dt} + \frac{\partial E}{\partial t}. \quad (1)$$

If the image brightness of each point in the image is supposed to be stationary with respect to the time variable (i.e., $dE/dt = 0$), then the following expression holds:

$$E_x u + E_y v + E_t = 0, \quad (2)$$

where the abbreviation for partial derivatives of the image brightness has been introduced, and u, v correspond to dx/dt, dy/dt, and represent the components of the local velocity vector $\mathbf{V}$ along the x and y directions, respectively.

A more general motion constraint equation was reported by Schunck in [12] and called Extended Optical Flow Constraint (EOFC) in [10]:

$$\nabla \cdot (E\mathbf{V}) + \frac{\partial E}{\partial t} = 0, \quad (3)$$

where:

$$\nabla \cdot (E\mathbf{V}) = \mathbf{V} \cdot \nabla E + E\nabla \cdot \mathbf{V}. \quad (4)$$

Equation (3) can be rewritten as:

$$E_x u + E_y v + E u_x + E v_y + E_t = 0. \quad (5)$$

A derivation for the EOFC was presented by Nagel in [2] choosing the same path as Schunck [13], [14]. Nagel's derivation was based on the fact that the local image irradiance can be considered, in first approximation, as the density in the image brightness features. The EOFC equation (5) differs from the OFC equation (2) only in the term involving the divergence of the optical flow field vector ($E\nabla \cdot \mathbf{V}$). If the EOFC is supposed to be the true expression of the optical flow field, OFC can be considered valid only in the region where the divergence of the optical flow field is equal to zero.

In general, the optical flow estimation suffers from two main problems. The first consists in the presence of discontinuities in the local velocity, related to image brightness discontinuities which are originated by the presence noise, too crisp patterns on the moving object surfaces, occlusions between moving objects, and too fast object velocities with respect to the system of measure. Generally speaking, this difficulty can be overcome (or simply attenuated) convolving the image with a 2-D or 3-D Gaussian smoothing operator.

The second is the so-called "problem of aperture" which is also present in the human vision. This is related to the impossibility to recover univocally the direction of motion if the object is observed through an aperture which is smaller than the object size. In this context, the references of the object under observation (such as textures – e.g., patterns) are not enough to perceive the transversal component of the object motion, and only the component of apparent velocity which is parallel to ∇E can be detected.

In the literature, two main approaches for optical flow estimation can be identified: the *regularization-* and the *multiconstraint-based* approaches.

The *regularization-based* approaches consider optical flow estimation as an ill-posed problem. Solutions are obtained minimizing a functional where a smoothness constraint is appropriately weighted to regularize the solution. The functional is minimized by using calculus of variations, and leads to define iterative solutions [1], [6], [15].

The *multiconstraint-based* approaches to optical flow estimation are based on the observation that the condition $dF/dt = 0$ can be made valid for any motion-invariant function F such as contrast, entropy, curvature, gradient magnitude, etc. By using a set of these constraints, which are evaluated at the same point in the image, a solvable system of equations, with u and v as unknowns, can be obtained [16], [17]. Other methods derive constraint equations by using the first and second derivatives of OFC or EOFC with respect to x, y and t [7], [8], [9], [10]. These multiconstraint-based approaches are solved with traditional numerical methods for the inversion or pseudo-inversion of the coefficient matrix.

3 Two Solutions for Optical Flow Estimation

Considering that the optical flow changes follow a law which is approximatively linear, a smoothed solution for the optical flow estimation can be derived from a linear approximation of the adopted constraint equation in the neighborhood of the point under consideration [18], [10]. This assumption is valid only if the optical flow field under observation is smooth. In this way, a set of similar constraints in the neighborhood of a pixel yields an over-determined system of equations. This approach is called "multipoint".

A multipoint solution based on the EOFC (5), as well as on the OFC (2), can be easily obtained. Numerical solutions can be obtained in the discrete domain, since images are sampled on a fixed grid of points at a regular time interval. Thus an image at time t is the collection of the irradiance measures $E_{(i,j,t)}$ for $i = 1,..M$, and $j = 1,..M$ along x- and y-axes, respectively. On this path, for the estimation of velocity components of the pixel under consideration an over-determined system of $N \times N$ constraint equations, where N is the dimension of the image segment side of the neighboring pixels, is defined in both cases.

In the case in which the constraint equation is the OFC, the system is formed by taking the following equations for all (i, j) in an $N \times N$ neighborhood of the estimation point:

$$E_{t(i,j,t)} + E_{x(i,j,t)} u + E_{y(i,j,t)} v = 0.$$

Since the equation is OFC, the over-determined system has 2 unknowns, and N must be grater than 2.

If the constraint is the EOFC, it should be noted that in the EOFC equation there are 2 unknowns (u_x, v_y) which are linearly dependent on each other since they have the same coefficient E. Thus the adopted constraint equation is:

$$E_t + E_x u + E_y v + E\nabla \cdot \mathbf{V} = 0, \quad (6)$$

and has been used to build an over-determined linear system of $N \times N$ equations (6) in 3 unknowns (u, v, $\nabla \cdot \mathbf{V}$) on $N \times N$ neighboring pixels (with $N \geq 2$). The system is formed by taking the following equations for all (i, j) in an $N \times N$ neighborhood of:

$$E_{t(i,j,t)} + E_{x(i,j,t)} u + E_{y(i,j,t)} v + E_{(i,j,t)} \nabla \cdot \mathbf{V} = 0,$$

Both the presented over-determined systems of equations have been solved by using a least-squares technique.

In the above techniques, a large N will had to smooth optical flow estimations. Furthermore, adopting large values for N leads to loss in resolution in the estimation of velocities. Therefore, the proposed methods can be used safely only when the velocity field is smooth in an $N \times N$ neighborhood, otherwise inaccurate results will be obtained.

Selection between OFC or EOFC equation mainly depend on the problem context (illumination sources, type of motion, surface reflectance...) [4]. Therefore, the choice should be made depending on the specific problem.

3.1 Computational complexity

Both multipoint solutions presented above can be regarded as a three phase process. The first phase addresses the estimation of constraint equations coefficients (e.g., E_x, E_y, E_t). The second phase regards the transformation of the over-determined system of $N \times N$ equations on n unknowns in an equivalent and determined system of n equations in n unknowns by using the least-squares technique. In the third phase the equivalent system of equations is solved to produce the optical flow components.

The explicit complexity for the presented multipoint solutions can be expressed by:

$$3M^2 + \left[\frac{M-2}{G}\right]^2 \left(N^2\left(\frac{n}{2}(n-1)+n\right)+n^3\right), \quad (7)$$

where: n is the number of the unknowns (2 for the OFC-based solution and 3 for the EOFC-based solution); M is the image dimension in pixels; G is the distance in pixels between two estimations point (i.e., the resolution); N is the dimension of the area side used for the estimation; and where the symbol $[x]$ is the greatest integer number smaller than x. The first term ($3M^2$), take into account the estimation of the partial derivatives E_x, E_y, and E_t of the image brightness; the second term is due to the least-squares technique, and to the method for solving the final system of equations with a LU decomposition (n^3).

The asymptotical complexity of (7) is:

$$O\left(\frac{M^2N^2n^2}{G^2}\right),$$

if $n < N^2$. This affirmation is always true since in our cases n is 2 or 3 (for the OFC- and EOFC-based solution, respectively) and N must be greater than 2 to define an over-determined system of equations.

If the dimension of the image area used for optical flow estimation, N, and the distance between spatially consecutive estimation point, G, are equals then the segments of estimation are tiled. On the contrary, if $G < N$ there is overlapping among the estimation areas which are close to each other.

The computational cost in term of number of floating point operations is reported in the following. This is useful in order to evaluate the computational power which is required to perform optical flow estimation in real-time – i.e., at video-rate frequency, 25 frames for second.

In total the number of floating point operations for the OFC-based solution ($n = 2$) is:

$$N_{ofc} = 6M^2 + \left[\frac{M-2}{G}\right]^2 (N^2 10 + 11),$$

and for the EOFC-based solution ($n = 3$) is:

$$N_{eofc} = 6M^2 + \left[\frac{M-2}{G}\right]^2 (N^2 18 + 30),$$

M	512	256	128	64	32
EOFC-based	126	32	8	2	0.49
OFC-based	70	17.5	4.3	1.1	0.27

Table 1: Millions of floating point operations (MFLOP), with $G = 1$, $N = 5$, for different image dimensions.

M	512	256	128	64	32
EOFC-based	3150	800	200	50	12.25
OFC-based	1750	437	107	27.5	6.7

Table 2: Millions of floating point operation for second (MFLOPS), with $G = 1$ and $N = 5$, for different image size.

In Tab.1 the number of floating point operations, expressed in millions, for the estimation of an optical flow field with $G = 1$, and $N = 5$ as a function of the dimension of the image is reported. In this case, $(M-2)^2$ velocity vectors at each time interval are estimated.

By observing Tab.2 can be seen that the computational requirements for the OFC-based solution are about 1/2 than the EOFC-based solution. These figures make possible to implement both these algorithms by using commercial processors with a floating point unit such as Motorola 96000, Intel i860, etc. provided that the image size is not large. In addition the number of MFLOPS should be divided at least by G^2 if it is not required to estimate a velocity vector for every pixel but only every G pixels, in both x and y direction.

The obtainable speed-up obtained by porting these algorithms on a parallel architecture with a single processor for each pixel is about M^2 (SIMD architecture). In this case, the power that is needed for each processor is very low. Under the hypothesis of $N = 5$, $G = 1$ in Tab.3 the number of FLOPS which must be executed for each PE to satisfy the real-time estimation constraint are reported.

It should be noted that according to the characteristics of the algorithm the number of FLOPS that each PE must execute are quite independent from the dimension of the image if the condition "one PE for each pixel is satisfied".

Algorithm	FLOPS/PE	μSec./FLOP
EOFC-based	12150	82
OFC-based	6675	149

Table 3: FLOPS/PE – Number of FLOPS which must be executed for each PE (FLOPS/PE) to estimate optical flow at real-time with $G = 1$ e $N = 5$. μSec./FLOP – execution time for a floating point operation on each PE of a SIMD parallel architecture.

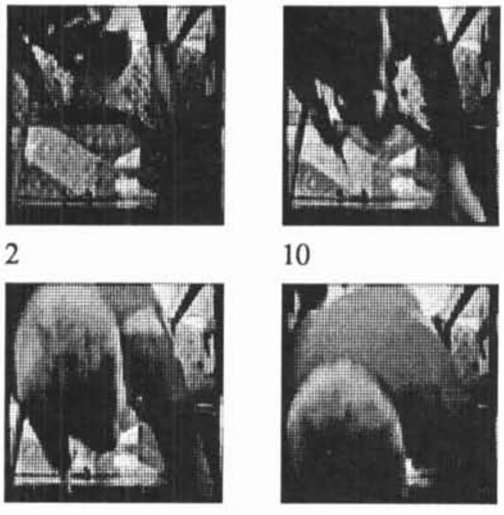


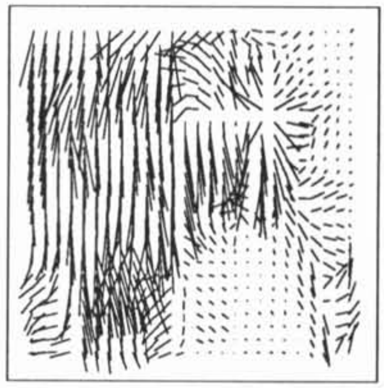
Figure 1: Sequence of images where a person is going into a bus (frames: 2, 10, 18, 26) (image resolution: 128×128).

4 Interpreting Optical Flow Fields in Time

The control of the number of people which are going into or out of some place is of interest in many applications. In museums, for example, it can be used to limit the number of people in certain areas. Controlling the number of people is also of interest for public services, as buses, in order to be able to properly schedule the frequence of service depending on the requirements. In many cases, to solve this problem mechanical systems, such as rotating tripod gates, and short iron doors. Unfortunately, these methods are not recommended where the velocity of people flowing in the entrance is high. This is what usually occur at the exit of many public places and when the people catch the bus where is not possible to place mechanical gates, since they produce a slowing down in the flow, which could be the cause of accidents.

The experimental results reported below refer to the case of counting people who are going into or out of a public bus. Two TV-cameras are placed just at the entrance of the bus over the stair steps, one camera for each lane, being the entrance divided by a metallic barrier. Images grabbed by the camera represent forms, belonging to the passing people, who are not completely in focus. In the sequence analysis the people have several distinct behaviors such as: entering or leaving the bus from the same door, stopping or swinging under the TV-camera, crowding conditions in which distinct people are very close each other under the TV-camera. A typical sequence presented in Fig.1 shows a person which is going into the bus.

Experiments reported in the following have been performed by using the EOFC-based solution. An optical flow field estimated by using the multipoint



18

Figure 2: Optical flow fields estimated from the sequence of images presented in Fig.1 by using the EOFC-based solution with $G = 1$ and $N = 5$ (frame: 18) .

EOFC-based solution on the sequence of Fig.1 is reported in Fig.2. The estimated optical flow shows that the identification of the shape of the moving person is not immediately detectable. This is mainly due to the fact that the body of the person is not completely in focus for the TV-camera. In addition the moving object is not rigid and its parts are moving with different velocities.

Problems that must be solved are: separation between distinct people who are close each other (this is what usually happened in the peak hours); recognizing the situations in which people remain for some reason motionless under the TV-camera; discerning from the people which are going into the bus with respect to those that are going out the bus.

In a sequence of optical flow estimations the velocity in a point i, j, at the time t is $\mathbf{V}_{i,j,t}$ with components $u_{i,j,t}$, and $v_{i,j,t}$. Considering that the flow of the people is only along the direction of the y-axis with respect to the image plane, the information related to the motion is only contained in the component $v_{i,j,t}$ of the optical flow $\mathbf{V}_{i,j,t}$. Thus, the spatio-temporal reasoning analyzes a moving object flow under the TV-camera only along the y-axis. A measure of flow can be obtained diving into s horizontal segments the optical flow field and estimating one velocity vector for each segment (see Fig.3):

$$v_{j,t} = \frac{1}{h} \sum_{i=1}^{h} v_{i,j,t},$$

for $j = 1...s$, where h is the number of optical flow vectors which are present in each row of the full optical flow field (i.e., $h = [(M-2)/G]^2$) (26 in the optical flow fields presented in Fig.2), and s is the number of optical flow vectors along the y-axis ($s = [(M-2)/G]^2$). In some cases, a smoother optical flow can be computed taking into account more that one row for each estimation of $v_{j,t}$.

To obtain more robust estimations in the presence of discontinuities due to noise and deformations, the

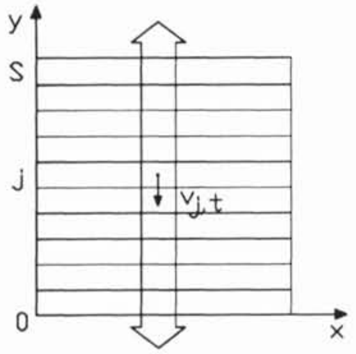


Figure 3: Optical flow field segments in the image plane at time t.

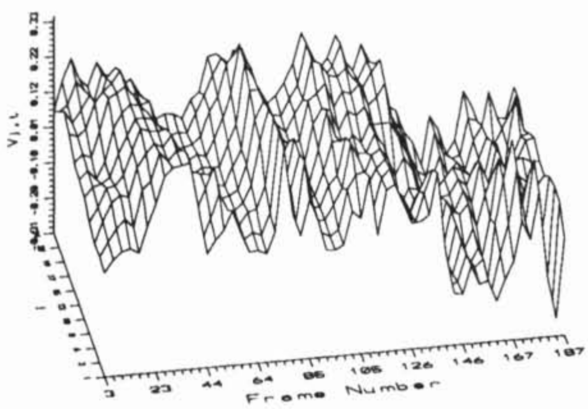


Figure 4: Spatio-temporal behavior of $\bar{v}_{j,t}$.

velocity estimated in each segment has been averaged also in time, by using three consecutive optical flow fields. Thus, the segment velocity is obtained as:

$$\bar{v}_{j,t} = \left(v_{j,t} + v_{j,t-1} + v_{j,t-2}\right)/3.$$

In Fig.4 the spatio-temporal behavior (frames 3 - 187) of the smoothed velocity component $\bar{v}_{j,t}$ along the y-axis is reported. In this surface it can be observed that the hill around the frame 30 corresponds to a person which is going into the bus. Analogously the negative hill around the frame 150 is due to a person that is going of out the bus. In Fig.5 a slice of the surface of Fig.4 is reported. From this curve the presence of at least three in-going and one outgoing passengers can be recognized, around the 30th, 71st, 110th, and 140th frame, respectively.

An important issue is to detect when the moving object enters or leaves the view area of the TV-camera. This can be useful to avoid the possibility of a wrong count when a person after being entered inside the area remains in the area motionless for several frames. To this end, it is useful to detect the boundaries of the optical flow fields. This can be done by using the derivative of the field with respect to the direction of motion. In this case the derivative of the smoothed velocity component $\bar{v}_{j,t}$ with respect to the y-axis can be estimated as:

$$\bar{v}_{j,t_y} = \left(\bar{v}_{j+1,t} - \bar{v}_{j,t}\right).$$

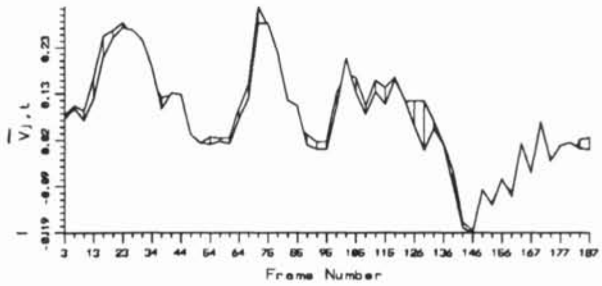


Figure 5: Temporal behavior of $\bar{v}_{j,t}$, for $j = 12$.

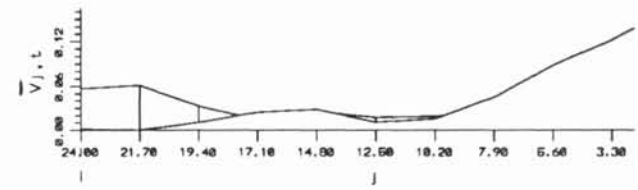


Figure 6: Motion profile for the frames 54-57, $\bar{v}_{j,t_{54}}$.

By using this information and the velocity of segments (for $j = 1...s$) is possible to detect the input/output of people in the view area and the direction of their motion. This mechanism can be seen, by observing the surface presented in Fig.4 at each frame time. In fact in this case the profile of the velocity of a moving person is depicted in Fig.6 for frames 54-57 of the sample sequence. This type of motion profile has been also used to solve the situations in which the people are strongly connected, since in these cases the pattern presents a minimun.

The computational cost for interpreting the optical flow fields in time in order to count the moving people is composed of the costs of: the estimation of the velocity vector for each segment from the entire optical flow field; the temporal averaging of the segment velocities, $\bar{v}_{j,t}$; the estimation of the derivatives, $\bar{v}_{i,t_y}$; and the reasoning to evaluate the counting. In general the trend of asymptotical complexity of interpretation is M^2/G^2. For this reason it can be neglected with respect to the complexity of optical flow estimation $O(M^2N^2n^2/G^2)$.

5 Conclusions

In this paper an optical flow technique has been proposed to solve the problem of counting people who are going in/out of controlled places. Computational requirements of two distinct solutions for optical flow estimation have been analyzed in order to verify the possibility of a real-time implementation on low-cost processors. Experimental results have been presented for the problem of counting people who get on/off a public bus.

References

[1] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17,

pp. 185–203, 1981.

[2] H.-H. Nagel, "On a constraint equation for the estimation of displacement rates in image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 13–30, January 1989.

[3] A. Verri, F. Girosi, and V. Torre, "Mathematical properties of the two-dimensional motion field: from singular points to motion parameters," *J. Opt. Soc. Am. A*, vol. 6, pp. 698–712, May 1989.

[4] P. Nesi, A. Del Bimbo, and J. L. C. Sanz, "A unified approach to optical flow constraint analysis," tech. rep., Dipartimento di Sistemi e Informatica, Facolta' di Ingegneria, Universita' di Firenze, DSI-RT 17/92, Italy, 1992.

[5] A. DelBimbo and P. Nesi, "Behavioral object recognition from multiple image frames," *Signal Processing*, vol. 27, no. 1, pp. 37–49, 1992.

[6] H.-H. Nagel, "Displacement vectors derived from second-order intensity variations in image sequences," *Computer Vision, Graphics, and Image Processing*, vol. 21, pp. 85–117, 1983.

[7] R. M. Haralick and J. S. Lee, "The facet approach to optical flow," in *Proc. of Image Understanding Workshop (Science Applications Arlington), Va* (L. S. Baumann, ed.), 1983.

[8] O. Tretiak and L. Pastor, "Velocity estimation from image sequences with second order differential operators," in *Proc. of 7th IEEE International Conference on Pattern Recognition*, pp. 16–19, 1984.

[9] A. Verri, F. Girosi, and V. Torre, "Differential techniques for optical flow," *J. Opt. Soc. Am. A*, vol. 7, pp. 912–922, May 1990.

[10] P. Nesi, A. DelBimbo, and J. L. Sanz, "Multiconstraints-based optical flow estimation and segmentation," in *International Workshop on Computer Architecture for Machine Perception*, (Paris), pp. 419–426, DGA/ETCA, CNRS/IEF and MEN/DRED, December 1991.

[11] A. Verri and T. Poggio, "Motion field and optical flow: Qualitative properties," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 490–498, May 1989.

[12] B. G. Schunck, "The motion constraints equation for optical flow," in *Proc. of 7th IEEE International Conference on Pattern Recognition*, pp. 20–22, 1984.

[13] B. G. Schunck, "Image flow: Fundamentals and future research," in *Proc. of IEEE Conference Computer Vision and Patter Recognition, CVPR'85*, (San Francisco, Ca, USA), pp. 560–571, 19-23 June 1985.

[14] B. G. Schunck, "Image flow continuity equations for motion and density," in *Proc. of Workshop Motion: Representation and Analysis*, (Charleston, SC, USA), pp. 89–94, 7-9 May 1985.

[15] E. C. Hildreth, "Computations underlying the measurement of visual motion," *Artificial Intelligence*, vol. 23, pp. 309–354, 1984.

[16] R. J. Woodham, "Multiple light source optical flow," in *Proc. of 3rd IEEE International Conference on Computer Vision ICCV'90, Osaka, Japan*, pp. 42–46, 4-7 Dec. 1990.

[17] A. Mitiche, Y. F. Wang, and J. K. Aggarwal, "Experiments in computing optical flow with the gradient-based multiconstraint method," *Pattern Recognition*, vol. 20, no. 2, pp. 173–179, 1987.

[18] M. Campani and A. Verri, "Computing optical flow from an overconstrained system of linear algebraic equations," in *Proc. of 3rd IEEE International Conference on Computer Vision ICCV'90, Osaka, Japan*, pp. 22–26, 4-7 Dec. 1990.