ROBUST ESTIMATION OF FOCUS OF EXPANSION AND DEPTH FROM HIGH CONFIDENCE OPTICAL FLOW

Sanbao Xu and Per-Erik Danielsson

Image Processing Laboratory, Department of Electrical Engineering
Linköping University, S-581 83 Linköping, Sweden
e-mail: xsb@isy.liu.se, ped@isy.liu.se

ABSTRACT

Camera translation is very useful for active perception of scene structure. Determining the camera motion, expressed as the Focus-Of-Expansion on the image plane, is important. A simple and robust method is presented for the estimation of FOE in translational motion, by the least squares technique using high confidence optical flow. Depth or time-to-collision is then recovered by two alternative methods using the FOE and optical flow or local image derivatives. Results with real image sequence are shown.

1. INTRODUCTION

Determining the camera motion is an important problem in visual motion where the observer interacts with the environment. The Focus of Expansion(FOE), which is the projection of the translation motion of the camera on the image plane, is an essential and very useful feature of the egomotion. Once located, among other things, it can be used to determine a dense flow field and to recover the depth or time-to-collision of visible surfaces.

Several methods have been suggested for computing the FOE in the literature. Extracting the FOE(or broadly, motion parameters) based on features points is suggested by Jain[1], Lawton[2], Ballard and Kimball[3], and Tsai and Huang[4]. Longuet-Higgins and Prazdny[5] compute motion parameters and scene structure from local optical flow and its derivatives. Negahdaripour and Horn[6] propose a direct method for locating the FOE, which exploits the positiveness of depth as a constraint without computing optical flow or establishing feature correspondence. Hummel and Sundareswaran[7,8] observe that the circular component of the flow field about (x_0, y_0) is a scalar function whose norm is quadratic in x_0 and y_0. They search the FOE at the minimum of this quadratic surface. Aloimonos[9] presents a voting technique for computing the direction of motion using the normal flow. Burger and Bhanu[10] compute the fuzzy FOE(a 2D region of possible FOE-locations), instead of looking for a single-point FOE.

In this paper we present a simple but robust method to estimate the FOE or camera motion from optical flow field computed from second derivatives in a sequence of images captured by a translating camera. It is well known that while camera rotation is very useful in tracking and fixation, it carries no information about scene structure. In contrast, camera translation is an essential way for active perception of scene structure through motion. Therefore, we feel that it is worthwhile to locate the FOE for pure translation, though it is a special case for rigid motion.

2. LEAST SQUARES SOLUTION TO THE FOE

For a translation motion (V_x, V_y, V_z), the true optic flow produced on the image plane(located at distance f) is given by

$$u = \frac{fV_x - xV_z}{Z}, \tag{1a}$$

$$v = \frac{fV_y - yV_z}{Z}. \tag{1b}$$

The focus of expansion(FOE) is defined as the point on the image plane at which the optical flow vanishes. That is,

$$(x_{FOE}, y_{FOE}) = (f\frac{V_x}{V_z}, f\frac{V_y}{V_z}). \tag{2}$$

Thus, each optic flow vector (u_i, v_i) satisfies the constraint

$$\frac{u}{v} = \frac{x - x_{FOE}}{y - y_{FOE}} \tag{3}$$

or

$$[x_{FOE}\ \ y_{FOE}]\begin{bmatrix} v_i \\ -u_i \end{bmatrix} = x_i v_i - y_i u_i. \tag{4}$$

By collecting many flow vectors in the image, we get a highly overdetermined linear system. The least squares method is then used to compute the FOE from the flow field,

$$[x_{FOE}\ \ y_{FOE}] = BA^T(AA^T)^{-1} \tag{5}$$

where

$$A = \begin{bmatrix} v_1 & v_2 & \dots & v_n \\ -u_1 & -u_2 & \dots & -u_n \end{bmatrix}, \tag{6}$$

$$B = [x_1v_1 - y_1u_1\ \ \dots .\ \ x_nv_n - y_nu_n]. \tag{7}$$

Explicitly, we have

$$x_{FOE} = \frac{\sum u_i^2 \sum v_i(x_i v_i - y_i u_i) - \sum u_i v_i \sum u_i(x_i v_i - y_i u_i)}{\sum u_i^2 \sum v_i^2 - (\sum u_i v_i)^2},$$

$$y_{FOE} = \frac{\sum u_i v_i \sum v_i(x_i v_i - y_i u_i) - \sum v_i^2 \sum u_i(x_i v_i - y_i u_i)}{\sum u_i^2 \sum v_i^2 - (\sum u_i v_i)^2}.$$

3. HIGH CONFIDENCE OPTICAL FLOW

It is now recognized in computer vision that confidence measure or uncertainty should be associated with various representations in vision systems(e.g. disparity in stereo). Optical flow is no exception. Determining the FOE by using the whole flow field would suffer from the ill–posedness of optical flow locally computed by differential methods. The optical flow computed by the first derivative method is actually the normal flow rather than the true flow(the so–called aperture problem). For the second derivative method, the computed optical flow is still not reliable in principally 1D variational(edge–like) image areas. It is known that the aperture problem disappears at corner–like brightness structure in the images. We use the second derivative method to compute optical flow, which provides a confidence measure associated with each flow vector. We have[11,12]

$$\begin{bmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_{xt} \\ I_{yt} \end{bmatrix} \tag{9}$$

or

$$\begin{bmatrix} u \\ v \end{bmatrix} = \frac{1}{G} \begin{bmatrix} I_{xt}I_{yy} - I_{yt}I_{xy} \\ I_{yt}I_{xx} - I_{xt}I_{xy} \end{bmatrix}, \tag{10}$$

where $G = I_{xy}^2 - I_{xx}I_{yy}$ is the determinant of the Hessian matrix of second partial derivatives, but also strongly connected to the Gaussian curvature of the image brightness surface. G is our confidence measure. High confidence indicates reliable optical flow such as those at 2D variational(hill shaped or saddle–like) image functions.

Therefore, we input the optical flow thresholded by the confidence measure G to the standard LS procedure. Alternatively, we can adopt the weighted LS method using $|G|$ *or* G^2 as the weighting factor.

4. COMPUTING THE DEPTH OR TIME–TO–COLLISION

Once the FOE is located with accuracy, we can computing the full flow by combining the FOE constraint with the Brightness Change Constant Equation(BCCE)

$$I_x u + I_y v + I_t = 0. \tag{11}$$

The full flow is given by

$$u = -\frac{I_t(x - x_{FOE})}{I_x(x - x_{FOE}) + I_y(y - y_{FOE})}, \tag{12a}$$

$$v = -\frac{I_t(y - y_{FOE})}{I_x(x - x_{FOE}) + I_y(y - y_{FOE})}. \tag{12b}$$

From the flow equations (1a) and (1b), we then recover the scene depth using the full flow

$$Z = V_z \frac{(x_{FOE} - x)u + (y_{FOE} - y)v}{u^2 + v^2}. \tag{13}$$

As an alternative we can use first order spatial and temporal image derivatives without actually computing the full flow,

$$Z = \frac{V_z}{I_t}\left[I_x(x - x_{FOE}) + I_y(y - y_{FOE})\right]. \tag{14}$$

More conveniently, in both (13) and (14) the relative depth or time–to–collision Z/V_z is determined by the FOE and the optical flow or image derivatives.

5. EXPERIMENTAL RESULTS

The FOE method is simple and efficient(LS, closed form), yet it is very robust since "better data is always better than more data". In this case, high confidence flow is better than the raw whole flow field. The robustness of the method is also reflected in the fact that for the determination of the FOE, only directional information, not the magnitude of the optic flow is needed. Experimental results with a real image sequence, the NASA sequence(the size, 512x512 pixels, of the original images is reduced to 256x256 pixels in a pyramid fashion), show that the error between the estimated FOE and the true FOE is within 3 ~ 4 pixels(Table 1 and Fig.1). In our experiments, the threshold of G is taken to be 4500 and the high confidence optical flows are located at the positions shown in Fig.2. As we see in Table 1, the G^2-weighted LS method gives slightly better performance. As the threshold of G is lowered, the weighted LS method is more evidently superior to the standard LS method(Table 2). The maps of the recovered relative depth or time–to–collision using the estimated FOE and (a) optical flow computed by first derivatives(eqs. 12 and 13), (b) optical flow computed by second derivatives(eqs. 10 and 13, and (c) image derivatives(eq. 14) are shown in Fig.3 (a), (b), and (c), respectively. Note that using second derivatives Fig.3c seems to give a more noise–free result. This is natural since we have used a noise–suppressing postprocessing[12] in computing optical flow by eq. (10). All the depth maps are filtered by a 3x3 median filter.

As seen from Tables 1 and 2, in the case of pure translation, the FOE can, indeed, be robustly estimated by the linear LS method from high confidence optical flow. Furthermore, the relative depth or time–to–collision is rather successfully computed via the shallow computations(in the sense of [13]) of eqs. (10), (12), and (13), or (14).

ACKNOWLEDGEMENTS

We are grateful for the grant no. 92–04827 supported by NUTEK(Swedish Board for Technical Development). The NASA image sequence is from the Database of the 1991 IEEE Workshop on Visual Motion at Sarnoff Research Center, Princeton.

REFERENCES

[1]. R. Jain, "Direct computation of the focus of expansion," IEEE Trans. Pattern Anal. Mach. Intell. PAMI-5, No.1 1983.

[2]. D. Lawton, "Processing translational motion sequences," CVGIP 22, 1983.

[3]. D. H. Ballard and O. A. Kimball, "Rigid body motion from depth and optical flow," CVGIP 22, No.1, 1983.

[4]. R. Y. Tsai and T. S. Huang, "Estimating three-dimensional motion parameters of rigid planar patch," Technical report R-922, Coordinated Science Lab., Univ. of Illinois, Urbana, 1981.

[5]. H. C. Longuet-Higgins and K. Prazdny, "The interpretation of a moving retinal images," Proc. Roy. Soc. London B 208, 1980.

[6]. S. Negahdaripour and B. K. P. Horn, "A direct method for locating the focus of expansion," CVGIP 46, No.3 1986.

[7]. R. Hummel and V. Sundareswaran, "Motion parameter estimation from global flow field data," IEEE Trans. Pattern Anal. Mach. Intell. to appear.

[8]. V. Sundareswaran, "A fast method to estimate sensor motion," Proc. of 2nd ECCV, St. Margherita, Italy, 1992.

[9]. Y. Aloimonos and Z. Duric, "Active egomotion estimation: A qualitative approch," Proc. of 2nd ECCV, St. Margherita, Italy, 1992.

[10]. W. Burger and B. Bhanu, "Estimating 3-D egomotion from perspective image sequences," IEEE Trans. Pattern Anal. Mach. Itell. PAMI 12, No.11, 1990.

[11]. S. Uras, F. Girosi, A. Verri, and V. Torre, "A computational approach to motion perception," Bio. Cyber. 60, 79-88, 1988.

[12]. P. E. Danielsson, P. Emanuelsson, et al., "Single chip high-speed computation of optical flow," IAPR Workshop on MVA'90, 331-335.

[13]. R. A. Brooks, A. M. Flynn and T. Marill,"Self calibration of motion and stereo vision for mobile robot navigation," Proc. DARPA Image Unders. Workshop, 1988, pp. 398-410.

Table 1. The estimated position of FOE from |G|>4500*

flow field	computed (xfoe, yfoe) LS	computed (xfoe, yfoe) weighted LS	flow number(percent) used
1	(125.91,117.94)	(125.72,116.98)	1088 (1.713%)
2	(127.22,116.38)	(127.16,116.10)	1088 (1.713%)
3	(127.59,118.04)	(127.26,117.42)	1101 (1.734%)
4	(127.97,117.71)	(127.27,117.37)	1131 (1.781%)
5	(127.03,117.85)	(126.88,117.38)	1122 (1.767%)

*The true FOE is at (124,116).

Table 2. The estimated position of FOE from |G|>1000*

flow field	computed (xfoe, yfoe) LS	computed (xfoe, yfoe) weighted LS	flow number(percent) used
1	(126.12,114.59)	(126.12,116.34)	6885 (10.84%)
2	(128.71,111.09)	(127.77,114.93)	6921 (10.90%)
3	(129.10,111.28)	(127.93,115.75)	6904 (10.87%)
4	(129.00,111.14)	(127.59,115.96)	6907 (10.88%)
5	(126.56,113.65)	(126.23,116.19)	6907 (10.88%)

*The true FOE is at (124,116).

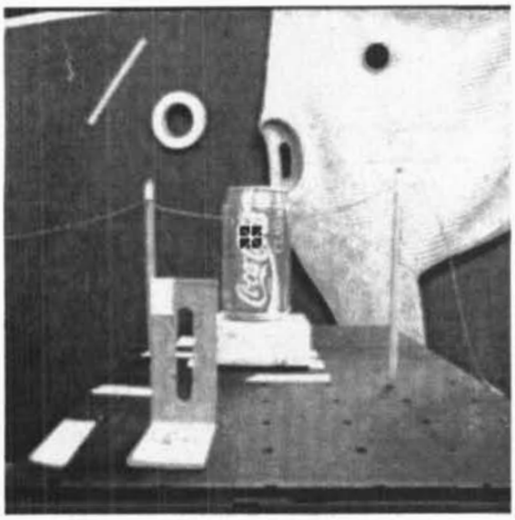


Fig.1 The computed and true FOE superimposed on an image of the sequence. Mark "+" stands for the true FOE and "x" for the estimated FOE.

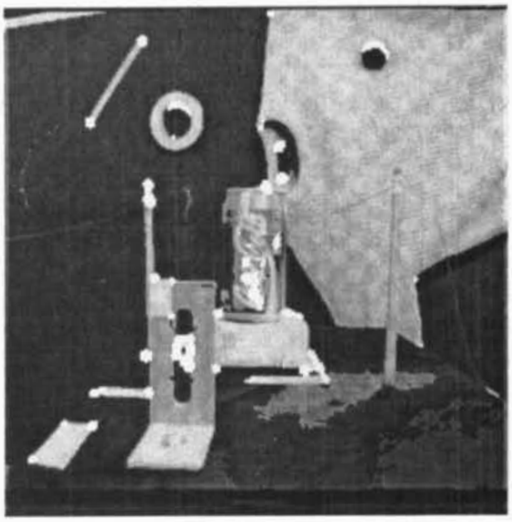
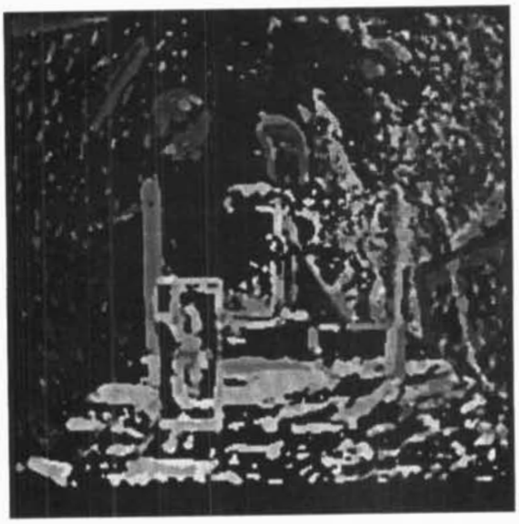
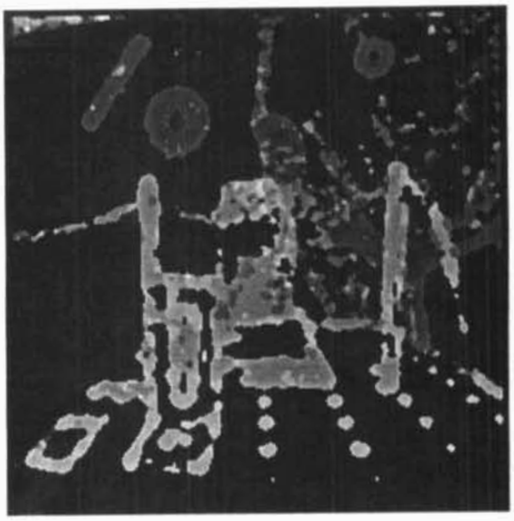


Fig.2 White areas: high confidence optical flow positions thresholded by G=4500.



(a)



(b)

(c)

Fig.3 The relative depth(or time-to-collision) maps computed from FOE and (a) optical flow by the first derivatives, (b) optical flow by the second derivatives, (c) first order image derivatives.