Presented at: 7th Int'l Conf Computer Analysis of Images and Patterns,
Kiel, Germany. September 10-12, 1997.

Discrete-Time Rigidity-Constrained Optical Flow*

Jeffrey Mendelsohn[1], Eero Simoncelli[2], and Ruzena Bajcsy[1]

[1] University of Pennsylvania, Philadelphia PA 19104-6228, USA
[2] New York University, New York NY 10003-6603, USA

Abstract. An algorithm for optical flow estimation is presented for the case of discrete-time motion of an uncalibrated camera through a rigid world. Unlike traditional optical flow approaches that impose smoothness constraints on the flow field, this algorithm assumes smoothness on the inverse depth map. The computation is based on differential measurements and estimates are computed within a multi-scale decomposition. Thus, the method is able to operate properly with large displacements (i.e., large velocities or low frame rates). Results are shown for a synthetic and a real sequence.

1 Introduction

Estimation of optical flow, a longstanding problem in computer vision, is particularly difficult when the displacements are large. Multi-scale algorithms can handle large image displacements and also improve overall accuracy of optical flow fields [1, 5, 9, 15]. However, these techniques typically make the unrealistic assumption that the flow field is smooth. In many situations, a more plausible assumption is that of a rigid world.

Given point and/or line correspondences, the discrete-time rigid motion problem has been studied and solved by a number of authors (e.g. [6, 7, 11, 14, 17]). For instantaneous representations, multi-scale estimation techniques have been used to couple the flow and motion estimation problems to provide a direct method for planar surfaces [4, 8]. These methods use the multi-scale technique to capture large motions while significantly constraining the flow with a global model. But the planar world assumption is quite restrictive, and the approach also contains a hidden contradiction; the algorithm can observe large image motions but can only represent small camera motions due to the instantaneous time assumption.

* J. Mendelsohn is supported by NSF grant GER93-55018. R. Bajcsy is supported by ARO grant DAAH04-96-1-0007, ARPA grant N00014-92-J-1647, and NSF grants IRI93-07126 and SBR89-20230. E. Simoncelli is supported by NSF CAREER grant MIP-9796040, the Sloan Center for Theoretical Neurobiology at NYU, and ARO/MURI grant DAAH04-96-1-0007 to UPenn/Stanford.

This paper describes an optical flow algorithm for discrete camera motion in a rigid world. The algorithm is based on differential image measurements and estimates are computed within a multi-scale decomposition; the estimates are propagated from coarse to fine scales. Unlike traditional coarse-to-fine approaches which impose smoothness on the flow field, this algorithm assumes smoothness of the inverse depth values.

2 Discrete-Time Optical Flow

The imaging system is assumed to use the following projection model:

$$\mathbf{p}_i = \mathbf{C}\mathbf{x}_i \frac{1}{z_i} \qquad \text{where} \quad \mathbf{x}_i \equiv \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} \quad \text{and} \quad \mathbf{p}_i \equiv \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} \ ; \tag{1}$$

$\mathbf{x}_i$ denotes the point's coordinates in the camera's frame of reference and $\mathbf{p}_i$ the image coordinates. The matrix $\mathbf{C}$ contains the camera calibration parameters and is presumed invertible.[3]

The discrete motion of a point is expressed as:

$$\mathbf{x'}_i = \mathbf{R}\mathbf{x}_i + \mathbf{t} \ , \tag{2}$$

where $\mathbf{R}$ is a (discrete-time) rotation matrix, $\mathbf{t}$ is a translation vector, and $\mathbf{x'}_i$ denotes the point's coordinates after the discrete motion.

A classic formulation of this constraint is due to Longuet-Higgins [10]:

$$\mathbf{x'}_i^T \left(\mathbf{t} \times \mathbf{R}\mathbf{x}_i\right) = 0 \ .$$

Using equation (1) to substitute for $\mathbf{x}_i$ gives:

$$z'_i \mathbf{p'}_i^T (\mathbf{C'}^{-1})^T \left(\mathbf{t} \times \mathbf{R}\mathbf{C}^{-1}\mathbf{p}_i z_i\right) = 0 \ . \tag{3}$$

Let $\mathbf{t}^\times$ represent the skew-symmetric matrix corresponding to a cross-product with $\mathbf{t}$. Using suitable linear algebraic identities, and assuming that $z_i \neq 0$ and $z'_i \neq 0$, leads to the following simplification:

$$\begin{aligned} 0 &= z'_i \mathbf{p'}_i^T (\mathbf{C'}^{-1})^T \mathbf{t}^\times \mathbf{R}\mathbf{C}^{-1}\mathbf{p}_i z_i \\ 0 &= \mathbf{p'}_i^T (\mathbf{C'}^{-1})^T \mathbf{t}^\times \mathbf{C'}^{-1}\mathbf{C'}\mathbf{R}\mathbf{C}^{-1}\mathbf{p}_i \\ 0 &= \mathbf{p'}_i^T (\mathbf{C'}\mathbf{t})^\times \mathbf{C'}\mathbf{R}\mathbf{C}^{-1}\mathbf{p}_i \\ 0 &= \mathbf{p'}_i^T \mathbf{L}\mathbf{p}_i \ , \end{aligned} \tag{4}$$

where $\mathbf{L} \equiv (\mathbf{C'}\mathbf{t})^\times \mathbf{C'}\mathbf{R}\mathbf{C}^{-1}$ is a matrix that depends on the global motion and camera calibration information. Equation (4) provides a constraint on the initial

[3] This assumption is valid for any reasonable camera system. For example, the pinhole model is included in this family, as well as more complex models such as that given in [16].

and final image positions assuming rigid-body motion and is, essentially, the fundamental matrix [12].

In addition, it will be useful to develop an expression for the final position, $\mathbf{p'}_i$, given the calibration, motion, and structure parameters. Substituting the inverse of equation (1) into the rigid-body motion constraint of equation (2):

$$\mathbf{C'}^{-1}\mathbf{p'}_i z'_i = \mathbf{R}\mathbf{C}^{-1}\mathbf{p}_i z_i + \mathbf{t} \ .$$

Solving for the image position after the motion:

$$\mathbf{p'}_i = \frac{\mathbf{C'}\mathbf{R}\mathbf{C}^{-1}\mathbf{p}_i + \mathbf{C'}\mathbf{t}\frac{1}{z_i}}{\hat{\mathbf{z}}^T\left(\mathbf{C'}\mathbf{R}\mathbf{C}^{-1}\mathbf{p}_i + \mathbf{C'}\mathbf{t}\frac{1}{z_i}\right)} \quad \text{where} \quad \hat{\mathbf{z}} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \ . \tag{5}$$

This rigid-world motion constraint must be connected with measurements of image displacements. Since differential optical flow techniques have proven to be quite robust [2], the formulation is based on the differential form of the 'brightness constancy constraint' [8]:

$$\left.\frac{\partial I}{\partial u}\right|_i \cdot \frac{\partial u_i}{\partial t} + \left.\frac{\partial I}{\partial v}\right|_i \cdot \frac{\partial v_i}{\partial t} + \left.\frac{\partial I}{\partial t}\right|_i = 0 \ .$$

Substituting discrete displacements for the differential changes in image positions[4] and rewriting to isolate $\mathbf{p}'_i$ gives:

$$\begin{bmatrix} \partial I/\partial u \\ \partial I/\partial v \\ \partial I/\partial t \end{bmatrix}_i^T \begin{bmatrix} u'_i - u_i \\ v'_i - v_i \\ 1 \end{bmatrix} = \begin{bmatrix} \partial I/\partial u \\ \partial I/\partial v \\ \partial I/\partial t \end{bmatrix}_i^T \begin{bmatrix} 1 & 0 & -u_i \\ 0 & 1 & -v_i \\ 0 & 0 & 1 \end{bmatrix} \mathbf{p}'_i = 0 \ . \tag{6}$$

This constraint is combined with equation (5), squared, and summed over a local neighborhood, $N(i)$, to produce an error metric:

$$E_i(\mathbf{A}, \mathbf{b}, \frac{1}{z_i}) = \frac{\left(\mathbf{A}\mathbf{p}_i + \mathbf{b}\frac{1}{z_i}\right)^T \mathbf{D}_i \left(\mathbf{A}\mathbf{p}_i + \mathbf{b}\frac{1}{z_i}\right)}{\left(\hat{\mathbf{z}}^T\mathbf{A}\mathbf{p}_i + \hat{\mathbf{z}}^T\mathbf{b}\frac{1}{z_i}\right)^2} \ ,$$

where $\mathbf{A} \equiv \mathbf{C'}\mathbf{R}\mathbf{C}^{-1}$, $\mathbf{b} \equiv \mathbf{C'}\mathbf{t}$, and $\mathbf{D}_i$ is a matrix constructed from the differential image measurements and known position vectors:

$$\mathbf{D}_i \equiv \begin{bmatrix} 1 & 0 & -u_i \\ 0 & 1 & -v_i \\ 0 & 0 & 1 \end{bmatrix}^T \left(\sum_{j \in N(i)} \begin{bmatrix} \partial I/\partial u \\ \partial I/\partial v \\ \partial I/\partial t \end{bmatrix}_j \begin{bmatrix} \partial I/\partial u \\ \partial I/\partial v \\ \partial I/\partial t \end{bmatrix}_j^T \right) \begin{bmatrix} 1 & 0 & -u_i \\ 0 & 1 & -v_i \\ 0 & 0 & 1 \end{bmatrix} \ ,$$

[4] The displacements are assumed to be small; this is reasonable given the multi-scale (coarse-to-fine) framework described in Section 3.

Minimizing $E_i(\mathbf{A}, \mathbf{b}, \frac{1}{z_i})$ with respect to $\frac{1}{z_i}$ gives:

$$\frac{1}{z_i} = -\frac{\mathbf{p}_i^T \mathbf{A}^T \mathbf{D}_i \left(\mathbf{I} - \mathbf{b}\hat{\mathbf{z}}^T\right) \mathbf{A}\mathbf{p}_i}{\mathbf{b}^T \mathbf{D}_i \left(\mathbf{I} - \mathbf{b}\hat{\mathbf{z}}^T\right) \mathbf{A}\mathbf{p}_i} \quad .$$

Substituting back into $E_i(\mathbf{A}, \mathbf{b}, \frac{1}{z_i})$, and noting that $\mathbf{L} = \mathbf{b}^\times \mathbf{A}$:

$$E_i(\mathbf{A}, \mathbf{b}) = -\frac{\mathbf{b}^T \mathbf{D}_i \left(\mathbf{L}\mathbf{p}_i\right)^\times \mathbf{D}_i \mathbf{A}\mathbf{p}_i}{\left(\mathbf{L}\mathbf{p}_i\right)^T (\hat{\mathbf{z}}^\times)^T \mathbf{D}_i \hat{\mathbf{z}}^\times \left(\mathbf{L}\mathbf{p}_i\right)} \quad .$$

Using a series of linear algebraic manipulations, the numerator may be rewritten as an expression quadratic in $\mathbf{L}$:

$$\begin{aligned} -(\mathbf{D}_i\mathbf{b})^T (\mathbf{L}\mathbf{p}_i)^\times \mathbf{D}_i \mathbf{A}\mathbf{p}_i &= (\mathbf{L}\mathbf{p}_i)^T (\mathbf{D}_i\mathbf{b})^\times \mathbf{D}_i \mathbf{A}\mathbf{p}_i \\ &= |\mathbf{D}_i| (\mathbf{L}\mathbf{p}_i)^T \mathbf{D}_i^{-1} \mathbf{b}^\times \mathbf{A}\mathbf{p}_i \\ &= |\mathbf{D}_i| (\mathbf{L}\mathbf{p}_i)^T \mathbf{D}_i^{-1} \mathbf{L}\mathbf{p}_i \quad , \end{aligned}$$

where $|\mathbf{D}_i|$ indicates the determinant of the matrix $\mathbf{D}_i$.

To efficiently solve for $\mathbf{L}$, a global metric is formed by summing over the image the weighted numerators of the $E_i(\mathbf{A}, \mathbf{b})$:

$$E(\mathbf{A}, \mathbf{b}) = \sum_i |\mathbf{D}_i| (\mathbf{L}\mathbf{p}_i)^T \mathbf{D}_i^{-1} \mathbf{L}\mathbf{p}_i / w_i \quad , \tag{7}$$

where w_i is the value of the denominator of $E_i(\mathbf{A}, \mathbf{b})$ using the previous estimate of $\mathbf{L}$. This metric is computed iteratively in the coarse-to-fine procedure and, from empirical observations, only one iteration at each scale is necessary.

The algorithm proceeds by first globally minimizing equation (7) to obtain a solution for the nine entries of the matrix $\mathbf{L}$, subject to the constraints $|\mathbf{L}| = 0$ and $\sum_j \sum_i (\mathbf{L}_{j,i})^2 = 1$ (to remove the scale ambiguity). Then, the squared optical flow constraint

$$E_f(\mathbf{p'}_i) = \mathbf{p'}_i^T \mathbf{D}_i \mathbf{p'}_i \quad , \tag{8}$$

is minimized at each pixel, subject to equation (4) with the estimated value of $\mathbf{L}$, to obtain an optical flow field.

3 Multi-Scale Implementation

Since the method is capable of estimating large (discrete) camera motions, it must be able to handle large image displacements. This is accomplished with a coarse-to-fine version of the algorithm on a multi-scale decomposition.

First, a Gaussian pyramid is constructed on the pair of input frames [3]. At the coarsest scale of the pyramid, the algorithm is employed as derived in the previous section to provide an initial coarse estimate of optical flow. This optical flow is interpolated to give a finer resolution flow field, denoted $(\Delta^c u_i, \Delta^c v_i)$. This

motion is removed from the finer scale images using warping; the warped images are denoted I^w.

Since the optical flow equation (8) is written only in terms of the final positions $\mathbf{p'}_i$, the constraint on the warped images needs only a slight modification:

$$\begin{bmatrix} \partial I^w/\partial u \\ \partial I^w/\partial v \\ \partial I^w/\partial t \end{bmatrix}_i^T \begin{bmatrix} 1 & 0 & -(u_i + \Delta^c u_i) \\ 0 & 1 & -(v_i + \Delta^c v_i) \\ 0 & 0 & 1 \end{bmatrix} \mathbf{p}'_i = 0 \ .$$

The remainder of the algorithm is as before: new matrices $\mathbf{D}_i$ are computed from this constraint and these are used to estimate $\mathbf{L}$ using equation (7). The weightings, w_i, are computed using the estimate of $\mathbf{L}$ from the previous scale. Finally, equation (8) is minimized at each pixel, subject to the constraint of equation (4), to estimate the optical flow.

4 Experimental Results

Experimental results were collected for three different algorithms on two sequences. The first method is a simple multi-scale optical flow (msof) algorithm [15]. The second computes flow for discrete motion of a planar world (planar) [13]. The third is the algorithm presented in this paper (rigid).

The first sequence is the 'Yosemite' sequence.[5] True optical flow vectors were computed from the motion, structure, and calibration data provided with the sequence. The textureless top region was ignored during error calculations. In order to obtain large motions, the computations are performed on the sequence subsampled at different temporal rates.

The second sequence was taken in the GRASP Laboratory from a camera mounted on a tripod. Six markers consisting of seven black disks on a white planar surface were placed in the scene and used to calculate ground-truth. Using knowledge of the individual targets, accurate centroids for the disks were computed. Flow was calculated from the motion of each centroid for a total of 42 flow vectors.

Table 1 shows results. The metric is the mean angular error in degrees [2]:

$$E_a = \frac{1}{n} \sum_{i=1}^{n} \cos^{-1}(\hat{\mathbf{v}}_{t_i}^T \hat{\mathbf{v}}_{e_i})$$

where $\hat{\mathbf{v}}_{t_i}$ is a unit three-vector in the direction of the true flow and $\hat{\mathbf{v}}_{e_i}$ is a unit three-vector in the direction of the estimated flow. The multi-scale optical flow algorithm did well for small motions but poorly for large ones. Since the range map for the Yosemite sequence is nearly planar, the difference in performance between the discrete algorithms is less significant than those for the real sequence. It is clear that the rigid algorithm provides the best optical flow estimates.

[5] This sequence was graphically rendered from an aerial photograph and range map by Lyn Quam at SRI.

sequence →	**Yosemite**							**Real**
frame interval →	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**1**
msof	6.13^o	7.49^o	13.10^o	21.32^o	29.09^o	35.82^o	42.43^o	3.28^o
planar	5.82^o	6.56^o	6.70^o	6.70^o	6.92^o	6.82^o	7.63^o	2.87^o
rigid	5.77^o	5.86^o	6.18^o	6.53^o	6.55^o	6.77^o	6.95^o	2.05^o

Table 1. Mean angular error in flow vectors for three different algorithms. See text.

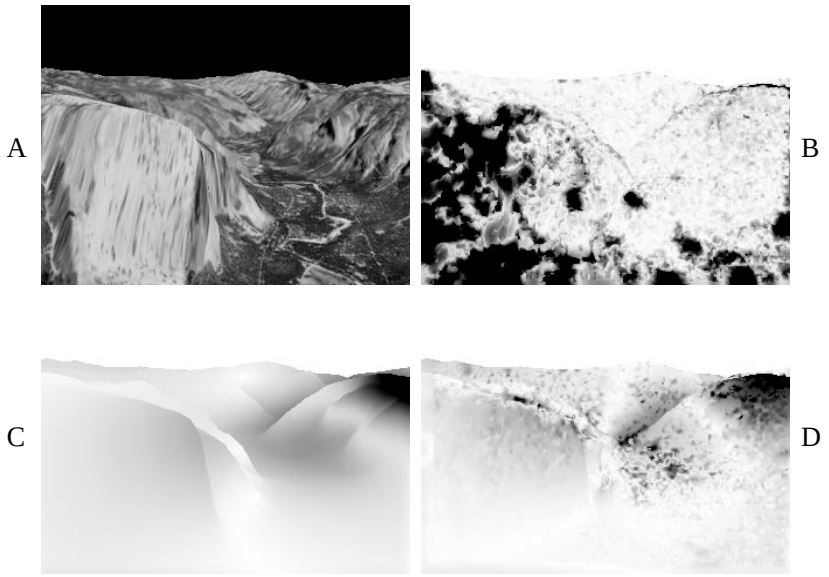

Fig. 1. Sample image from the Yosemite sequence (A) and the angular error metric of the computed flow fields for the Yosemite sequence by the msof (B), planar (C), and rigid (D) algorithms with a frame interval of four. White corresponds to 0^o of error and black to 45^o.

5 Conclusion

A multi-scale algorithm for estimating optical flow based on an uncalibrated camera moving through a rigid world has been presented. Its implementation is only slightly more complicated and time-consuming than standard multi-scale algorithms. In situations where the camera may be undergoing relatively large motions, the superiority of the rigid model has been demonstrated on both a synthetic and a real sequence.

References

1. P. P. Anandan. A Computational Framework and an Algorithm for the Measurement of Visual Motion. *IJCV*, 2, 283–310, 1989.

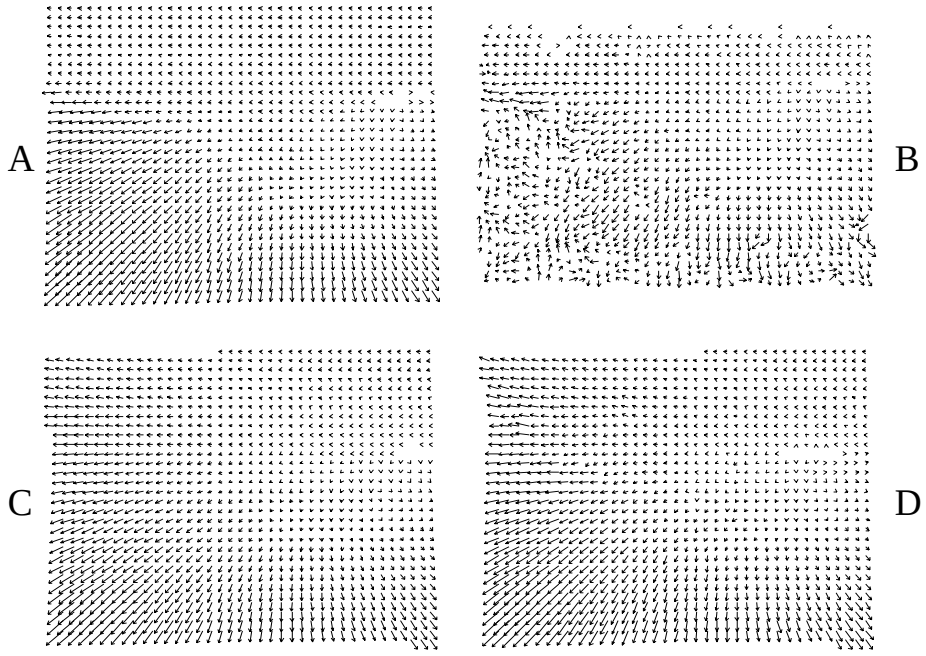

Fig. 2. True flow field (A) and computed flow fields for the Yosemite sequence by the msof (B), planar (C), and rigid (D) algorithms with a frame interval of four.

2. J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of Optical Flow Techniques. *IJCV*, 1992.
3. P. J. Burt. Fast filter transforms for image processing. *Computer Graphics and Image Processing*, 16, 20–51, 1981.
4. J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. Hierarchical Model-Based Motion Estimation. *Proc. ECCV*, 237–252, Springer-Verlag, Santa Margherita Ligure, Italy, 1992.
5. W. Enkelmann and H. Nagel. Investigation of Multigrid Algorithms for Estimation of Optical Flow Fields in Image Sequences. *Computer Vision Graphics Image Processing*, 43, 150–177, 1988.
6. R. Hartley. Projective Reconstruction from Line Correspondences. *Proc. IEEE CVPR*, 1994.
7. R. Hartley, R. Gupta, and T. Chang. Stereo from Uncalibrated Cameras. *Proc. IEEE CVPR*, 1992.
8. B.K.P. Horn. *Robot Vision*. MIT Press, Cambridge, MA, 1986.
9. M. Leuttgen, W. Karl, and A. Willsky. Efficient Multiscale Regularization with Applications to the Computation of Optical Flow. *IEEE Trans. Im. Proc.*, 1992.
10. H. C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. *Proc. Royal Society of London B*, 208, 385–397, 1980.

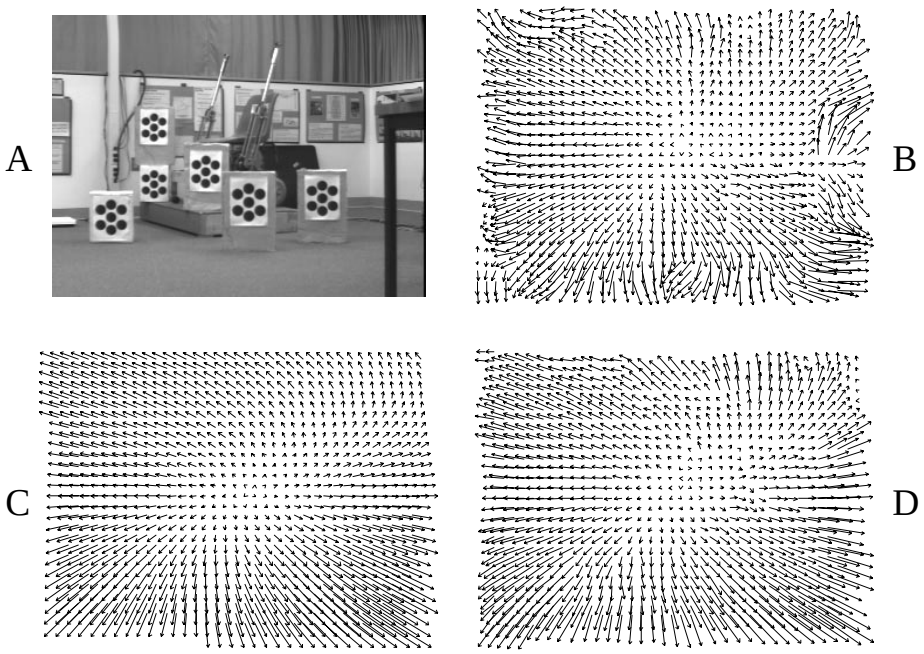

Fig. 3. Sample image from the real sequence (A) and the computed flow fields for the real sequence by the msof (B), planar (C), and rigid (D) algorithms.

11. Q.T. Luong and T. Vieville. Canonic Representation for the Geometry of Multiple Projective Views. *Proc. ECCV*, Stockholm, 1994.
12. Q.T. Luong and O. Faugeras. The Fundamental Matrix: Theory, Algorithms, and Stability Analysis. *IJCV*, 43–76, 1996.
13. J. Mendelsohn, E. Simoncelli, and R. Bajcsy. Discrete-Time Rigid Motion Constrained Optical Flow Assuming Planar Structure. *Univ. of PA GRASP Laboratory Technical Report #410*, 1997.
14. A. Shashua and N. Navab. Relative Affine Structure: Theory and Application to 3D Reconstruction from Perspective Views. *Proc. IEEE CVPR*, Seattle, June 1994.
15. E. P. Simoncelli, E. H. Adelson, and D. J. Heeger. Probability Distributions of Optical Flow. *Proc. IEEE CVPR*, Maui, June 1991.
16. T. Vieville and O. D. Faugeras. The First Order Expansion of Motion Equations in the Uncalibrated Case. *Computer Vision and Image Understanding*, 128–146, July 1996.
17. T. Vieville, C. Zeller, and L. Robert. Using Collineations to Compute Motion and Structure in an Uncalibrated Image Sequence. *IJCV*, 1995.

This article was processed using the LaTeX macro package with LLNCS style.