Removal of Translation Bias when using Subspace Methods

W. James MacLean
Department of Computer Science,
University of Toronto,
Toronto, Ontario M5S 3H5
maclean@vis.toronto.edu

Abstract

Given estimates of the motion field (optic flow) from an image sequence, it is possible to recover translational direction, $\vec{T}$, using a variety of techniques. One such technique, known as "subspace methods," generates constraints which are perpendicular to $\vec{T}$, so that two distinct constraints allow a solution for $\vec{T}$. In practice many constraints are used in a least-squares solution, but it has been observed that the recovered estimates for $\vec{T}$ are biased towards the optical axis. While the cause of the bias is well known, previous attempts to remove it have been flawed. This paper outlines a new method which removes the bias. The technique is simple to apply and computationally efficient.

1 Introduction

In the analysis of image sequences in which the observer moves relative to a static background, the recovery of the observer's motion parameters (termed *egomotion*) has long been a central problem. A variety of methods have been introduced to accomplish this: methods have been introduced by Bruss and Horn [1], Rieger & Lawton [16], Longuet-Higgins & Prazdny [14, 19], and Jepson & Heeger [7, 9, 10, 8, 11].

The subspace methods allow one to recover translational motion $\vec{T}$ using a linear method [11]. However, it is immediately apparent that for the case of noisy flow as input, a significant bias in the estimate for $\vec{T}$ is seen. This occurs as a result of the fact that isotropic noise in flow measurements used as input lead to anisotropic noise in constraints which are central to the subspace methods. The bias is consistently towards the optic axis.

Two different approaches have been used to compensate for the bias. Jepson & Heeger [11] suggest a *dithering* method in which more noise is added to the constraints with the goal of making the resulting noise isotropic in nature. Kanatani [12, 13] takes the approach of subtracting the anisotropic covariance matrix prior to estimating $\vec{T}$, in a process term *renormalization*. Since subtraction is being used it is necessary to correctly estimate the scale of the subtracted covariance matrix.[1]

In this paper the author presents a new method for dealing with the bias. This method involves rescaling the linear constraints from the subspace methods according to the covariance matrix for the constraints. Since the re-scaling is done through a multiplication, we eliminate the need to estimate the scale of the covariance matrix.

In this paper the theory of subspace methods and the cause of the bias are presented, the rescaling method for bias removal is introduced, and results from a synthetic image sequence (for which the "ground-truth" motion are known) are presented.

2 Theory

This paper considers the case of planar receptor surfaces. The coordinate system (see Figure 1) is right-handed and aligned so that the z-axis aligns with the optical axis of the camera. The origin of the coordinate system is placed at the optical centre of the camera, so the receptor surface is the plane $z = f$. The receptor surface is placed in front of the optic centre to avoid the need to reflect coordinates.

A point $\vec{X}$ in 3-D space images to point $\vec{x}$ in the image place. Under perspective projection we have

$$\vec{x} = \frac{f}{X_3}\vec{X} \; . \tag{1}$$

We are interested in the movement of $\vec{x}$, which has been termed the *motion field*. Assuming

$$\vec{V} = \frac{d\vec{X}}{dt} = \vec{T} + \vec{\Omega} \times \vec{X}$$

where $\vec{T}$ is the translational motion and $\vec{\Omega}$ the rotational motion of $\vec{X}$, then we can write

$$\vec{u} = \frac{d\vec{x}}{dt} = \begin{bmatrix} 1 & 0 & -x_1/f \\ 0 & 1 & -x_2/f \\ 0 & 0 & 0 \end{bmatrix} \left(\frac{f}{X_3}\vec{T} + \vec{\Omega} \times \vec{x} \right) \; . \tag{2}$$

The vector $\vec{u}$ represents the motion of $\vec{x}$ in the image plane. The measurement of motion of image points is often referred to as *optic flow*.

[1] Kanatani's analysis is with respect to constraints on $\vec{T}$ derived using the essential matrix methods, so the bias problem is not exclusive to the subspace methods.

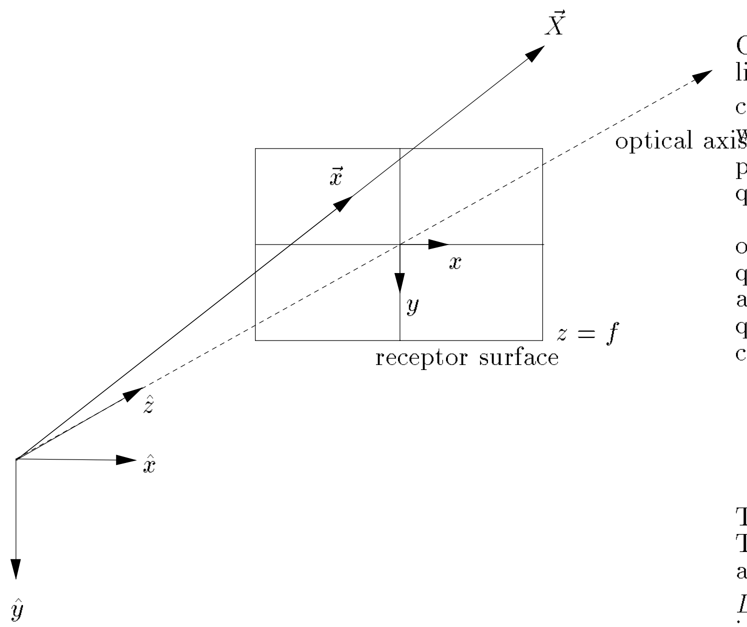

Figure 1: A right-handed coordinate system is attached to the imaging system. The origin coincides with the nodal point of the imaging system, and the z-axis with the optical axis. The planar receptor surface lies in the $z = f$ plane. A point $\vec{X}$ in the 3-D world is imaged to a point $\vec{x}$ in the image plane. Under perspective projection the relation is $\vec{x} = \frac{f}{\vec{X}^T \hat{z}} \vec{X}$.

Knowledge of the motion field can be used to estimate the values of $\vec{T}$ and $\vec{\Omega}$ if the motion field arises from a single underlying motion.[2] Techniques for recovering motion parameters have included methods based on orthographic projection [18], the essential matrix [14, 19], methods which require the detection of planar surfaces in the image [17, 2], and *subspace methods*. The latter probably begins with work by Rieger & Lawton [16] and was further developed by Heeger & Jepson [7, 9, 10, 8, 11].

A simple algebraic manipulation of Eq. 2 leads to the following *bilinear constraint* on translation and rotation:

$$\vec{T}^T (\vec{x} \times \vec{u}) + (\vec{T} \times \vec{x})^T (\vec{x} \times \vec{\Omega}) = 0 \ . \qquad (3)$$

One such constraint can be written for each $\vec{u}$. Nonlinear methods must be applied to a set of bilinear constraints to recover the values of $\vec{T}$ and $\vec{\Omega}$. It is worth noting that Eq. 3 can be rewritten more simply as $\vec{T}^T (\vec{a} + B\vec{\Omega})$ where B is a 3×3 matrix-valued quadratic function of image position $\vec{x}$ and $\vec{a} = \vec{x} \times \vec{u}$.

A *linear constraint* can be constructed from 7 or more bilinear constraints. Noting that B is a quadratic value function, it is possible to compute a set of coefficients $\{c_k\}$ which are orthogonal to all quadratic forms in $\vec{x}$, and when applied to the bilinear constraint effectively annihilate the B terms, leaving

$$w\vec{\tau} = \sum_{k=1}^{K} c_k [\vec{u}(\vec{x}_k) \times \vec{x}_k] \ . \qquad (4)$$

The unit-vector $\vec{\tau}$ is guaranteed to be orthogonal to $\vec{T}$. Thus 2 or more linear constraints can be used to obtain a linear solution for $\vec{T}$. Specifically, if we construct $D = \sum_{i=1}^{N} w_i^2 \vec{\tau}_i \vec{\tau}_i^T$ then the eigenvector corresponding to the minimum eigenvalue of D is the recovered direction for $\vec{T}$.[3]

In the absence of noise the linear subspace constraints are exact [11]. Unfortunately optic flow estimates tend to be very noisy, so it is necessary to consider the effect this has on the resulting $\vec{\tau}$ vectors. Jepson & Heeger [11] report a bias in the estimates of $\vec{T}$ when using a motion field with isotropic noise added to generate the linear constraints. The cause of the bias lies in the fact that the noise in the linear constraints is anisotropic.

Consider the noisy constraint vector $\tilde{\tau} = \vec{\tau} + \vec{n}$ where $E\{\vec{n}\} = 0$, and

$$E\{\vec{n}\vec{n}^T\} = \sigma^2 \begin{bmatrix} 1 & 0 & -x \\ 0 & 1 & -y \\ -x & -y & x^2 + y^2 \end{bmatrix} \qquad (5)$$

[2] There are techniques for recovering motion parameters from a motion field which arises from multiple underlying motions, but these are not central to the point of this paper, and as such are not discussed here.

[3] Note that we cannot recover the magnitude of $\vec{T}$, just its direction. This is inherent to the problem itself and not just in the particular case of subspace methods.

where (x, y) is the average image location of the bilinear constraints used to construct $\vec{\tau}$. This covariance matrix for the noise vectors is derived assuming isotropic, zero-mean noise in the optic flow. Recalling the construction of our D matrix,

$$\begin{aligned}\tilde{D} &= \sum_{i=1}^{N} w_i^2 \tilde{\tau}_i \tilde{\tau}_i^T \\ &= \sum_{i=1}^{T} w_i^2 (\vec{\tau}_i \vec{\tau}_i^T + \vec{\tau}_i \vec{n}_i^T + \vec{n}_i \vec{\tau}_i^T + \vec{n}_i \vec{n}_i^T) \ ,\end{aligned}$$

we see that the expected value for $\tilde{D}$ becomes

$$\begin{aligned}E\{\tilde{D}\} &= D + E\{\sum_{i=1}^{N} w_i^2 \vec{n}_i \vec{n}_i^T\} \\ &= D + \sigma^2 M\end{aligned}$$

where

$$M = \sum_{i=1}^{N} w_i^2 \begin{bmatrix} 1 & 0 & -x_i \\ 0 & 1 & -y_i \\ -x_i & -y_i & x_i^2 + y_i^2 \end{bmatrix} .$$

We see that the noise adds a term to the expected value of $\tilde{D}$, and as such we expect it to affect the eigenvectors of $\tilde{D}$.

An intuitive explanation as to why the estimate for $\vec{T}$ is biased towards the centre of the image is as follows. While the $\vec{\tau}$ are orthogonal to $\vec{T}$, we see that for small angular extent of the imaging receptor, the constraints are also roughly orthogonal to the optic axis, since the $\vec{\tau}$ constraints are constructed as the sume of terms involving $\vec{u}_i \times \vec{x}_i$ and the $\vec{x}_i$ used to construct each constraint do not vary largely. In this case translational direction estimates near the optical axis are 'favoured'. We expect that as angular extent of the image is decreased that the bias will become worse, and this is indeed what happens[11].

Jepson & Heeger [11] suggested a dithering method to make the noise in the $\vec{\tau}_i$ isotropic, thereby removing the bias. While this approach is effective, it is not an intuitively satisfying approach since it involves adding more noise. Note that the bilinear constraints themselves do not suffer from the bias, so another approach is to get an initial (biased) estimate for $\vec{T}$ using the linear constraints, and then improve this estimate through a small number of iterations to solve the bilinear constraints.

Kanatani [13] suggests a method called *renormalization* for removal of the bias. Recalling that $E\{\tilde{D}\} = D + \sigma^2 M$, it is possible to construct $\hat{D} = \tilde{D} - \sigma^2 M$. This leads to $E\{\hat{D}\} = D$. This however requires an estimate not only of M but of its scaling factor σ^2, which is related to the noise in the optic flow. By contrast, the proposed re-scaling method does not require an estimate for σ^2.

2.1 Removing the Bias

We have derived the form of the noise covariance matrix for the linear constraints. In general we will not know the scaling of the noise, *i.e.* we won't know the value of σ, so it is not feasible to subtract $\sigma^2 M$ from $\tilde{D}$, as suggested by Kanatani [13]. However, it is possible to *re-scale* the $\vec{\tau}_i$ into a space where the noise is isotropic. If $\vec{T}$ is estimated in this re-scaled space the bias will have been removed. Finally, the estimate for $\vec{T}$ can be converted back to the original space.

In order to understand how this works, note that adding a scaled version of the identity matrix to D does not change the eigenvectors: if $\tilde{D} = D + \sigma^2 I_3$, then $D\vec{x} = \lambda\vec{x} \rightarrow \tilde{D}\vec{x} = (\lambda + \sigma^2)\vec{x}$. The eigenvectors of the two matrices are identical, and the ordering of the eigenvalues is preserved. Re-scaling is achieved by pre- and post-multiplying $\tilde{D}$ by the inverse square-root of the covariance matrix M. This gives us $M^{-1/2} D M^{-1/2} + \sigma^2 I_3$, which has the same eigenvectors as $\hat{D} = M^{-1/2} D M^{-1/2}$. This operation is sometimes called *pre-whitening* [5]. Choose the eigenvector $\vec{x}$ that corresponds to the minimum eigenvalue of $\hat{D}$, namely $M^{-1/2} D M^{-1/2} \vec{x} = \lambda\vec{x}$. The new estimate for the translational direction is $\vec{T} = M^{-1/2}\vec{x}$.

Note that $M^{-1} D (M^{-1/2}\vec{x}) = \lambda M^{-1/2}\vec{x}$. This means the estimate for $\vec{T}$ is an eigenvector of $M^{-1}D$, not D. However, since D represents the noise-free constraints, its minimum eigenvalue is 0. This guarantees that pre-multiplying D by M^{-1} will not change the corresponding eigenvector. Therefore the estimate $\vec{T}$ corresponds to the noise-free estimate, and the bias has been removed. It can be shown that this is also the maximum-likelihood (ML) estimate. It will not be possible to completely remove the bias in practice, as the form M depends on isotropic noise in the flow estimates, and will not be exact for any case of anisotropy in the noise.

It is worth noting that once $M^{-1/2}$ has been computed, its application involves three matrix multiplications, and therefore is computationally inexpensive. Computing M can be done in advance if the constraint weights are assumed to be equal.

3 Results

The rescaling technique developed in the previous section is now applied to a synthetic image sequence. The first task is to estimate the covariance form matrix, M, which will be used in the re-scaling step.

One possibility is to compute the "average image location" of the constraints, $\vec{x}_{av}$, and compute M as

$$M_{av} = \begin{bmatrix} 1 & 0 & -x_{av} \\ 0 & 1 & -y_{av} \\ -x_{av} & -y_{av} & x_{av}^2 + y_{av}^2 \end{bmatrix} .$$

It might be argued that this will always be at the centre of the image, but since many $\vec{\tau}$ constraints will be rejected due to low SNR, it in general will not be. It would also be possible to weight the constraint location vectors by each constraint's SNR in order to get a

Figure 2: On top is a depth-map (Z-buffer) from a computer generated image of an office. Below is a depth-map for a cube. These two depth-maps have been used to generate a synthetic flow field containing an independently moving object.

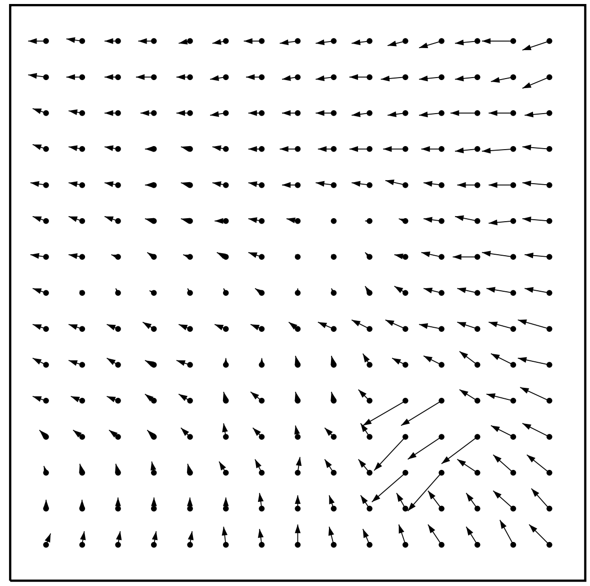

Figure 3: This figure shows synthetic optic flow generated from the depth maps shown in Figure 2. The observer is moving with a translational velocity of $\vec{T} = [1\ 0\ 1]^T$ with respect to the background. The cube is falling, and has a translational velocity of $\vec{T} = [0\ 1\ 0]^T$. A rotation has been added to simulate the observer fixating a point near the centre of the image. 10% noise has been added.

"centre of mass" type of average image position. This is referred to as 'Method 1.'

A second method would be to compute M as the average of the M_i for each constraint, where M_i is given by Eq 5. This method will be referred to as 'Method 2.' Again, this average could be weighted by the SNR of individual constraints. Results from both methods are demonstrated.

Motion field estimates were generated by applying a known $\vec{T}$ and $\vec{\Omega}$ to the depth-map (Z-buffer) for the synthetic image (see Figure 2). The translational and rotational motions were chosen such that the back of the chair was "fixated". This has the effect of improving SNR for the resulting linear constraints [15]. The depth map is shown in Figure 2, and the resulting (noisy) flow is shown in Figure 3.[4]

Random noise was added to the optic flow in a series of 5 trials. The noise is added as $\hat{u}(\vec{x}) = \vec{u}(\vec{x}) + \vec{n}$. The noise component $\vec{n}$ is chosen from a 2-D isotropic normal distribution having a standard deviation equal to 10% of $||\vec{u}(\vec{x})||$. The noisy flow field is shown in Figure 3. Note that this noise is multiplicative in nature, but this has been suggested as an appropriate model for optic flow recovery [4, 6]. Linear constraints were computed from the flow using a 7×7 convolution mask

[4]The flow and motion estimates were generated as part of a project on segmenting multiple 3-D motions from optic flow. Results shown in this paper are for the background motion only.

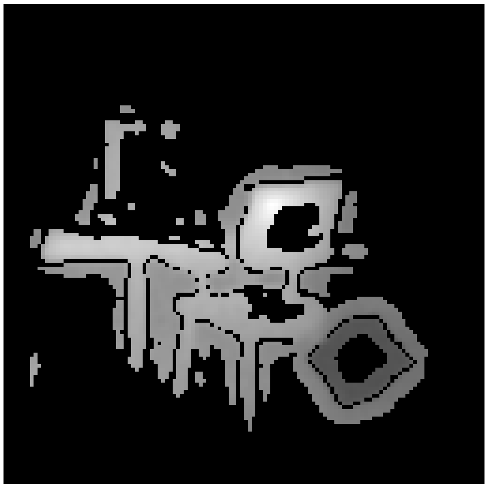

Figure 4: This is a plot of the magnitudes of the $\vec{\tau}$'s recovered from Figure 3. Regions containing depth-discontinuities give rise to the largest constraints. Constraints having an SNR of less than 5 were removed.

	True	Uncorrected	Method 1	Method 2
$\vec{T}$	$\begin{bmatrix} 0.7071 \\ 0.0000 \\ 0.7071 \end{bmatrix}$	$\begin{bmatrix} 0.6316 \\ 0.0005 \\ 0.7737 \end{bmatrix}$	$\begin{bmatrix} 0.7236 \\ 0.0037 \\ 0.6890 \end{bmatrix}$	$\begin{bmatrix} 0.7196 \\ 0.0035 \\ 0.6932 \end{bmatrix}$
error	0.0°	5.7523°	1.4172°	1.0907°

Table 1: This table shows the results of correcting for the anisotropic nature of the noise on the estimated translational direction. The results are tabulated over 5 trials, each of which uses a different seed to the random number generator to add noise to the optic flow. Both Method 1 and Method 2 provide considerable improvement over the uncorrected case.

[11] (see Figure 4) and estimates for $\vec{T}$ were computed both with and without the re-scaling method. The results are shown in Table 1.

The "true" (noise-free) direction for $\vec{T}$ is $[0.7071\ 0\ 0.7071]^T$. This corresponds to translation which is forward and to the right. The estimate of $\vec{T}$ recovered without bias- correction has an error of 5.7523°, and is biased towards the centre of the image as expected.

From Table 1 we see that both Method 1 and Method 2 offer substantial improvement over the uncorrected solution, with Method 2 performing somewhat better. The exact form used for C is therefore important, and should be made the subject of further study. Note that the value of M used in both methods is only an approximation to the form we require, so it is not expected that the bias will be completely removed.

4 Conclusion

Estimation of egomotion parameters from image sequences is an important pursuit in computer vision. The development of linear methods for estimating translational direction is an important step towards fast estimation, but comes with an inherent problem in the form of biased estimates.

While a number of methods have been proposed for solving this bias problem, they require either adding more noise, which is intuitively un-appealing, or attempting to estimate the scale of the covariance matrix, which is difficult unless *a priori* information about the amount of noise in the flow estimates is available.

The author has presented a new method for removing the bias which only requires knowledge of the form of the covariance matrix, that is knowledge of the covariance matrix up to a scale factor. Removal of the requirement that the scale is known is a significant improvement, since this parameter is difficult to estimate.

Further results from the re-scaling method presented in this paper may be found in [3] and the method is shown to outperform both dithering and renormalization. Earnshaw & Blostein [3] suggest a further improvement which requires an iterative solution.

Results are presented for optic flow from a synthetic image sequence. Use of a synthetic sequence allows knowledge of the "ground truth" egomotion parameters, demonstration of the bias effect, and demonstration of the efficacy of the technique for its removal. Two different methods are suggested for estimating the covariance matrix (up to a scale factor) associated with the subspace linear constraints. Since the estimates for the form of the covariance matrix are not exact the bias was not completely removed, but was significantly reduced.

5 Acknowledgements

The author would like to thank Allan Jepson, who made many suggestions and contributions to this work, and to NSERC for providing financial support.

References

[1] Anna R. Bruss and Berthold K. P. Horn. Passive navigation. *Computer Vision, Graphics and Image Processing*, 21:3–20, 1983.

[2] Trevor Darell and Alexander Pentland. Robust estimation of a multi-layered motion representation. In *Proceedings of the IEEE Workshop on Visual Motion*, pages 173–177, Princeton, New Jersey, October 1991.

[3] A. Mark Earnshaw and Steven D. Blostein. The performance of camera translation direction estimators from optical flow: Analysis, comparison, and theoretical limits. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):927–932, 1996.

[4] David J. Fleet. *Measurement of Image Velocity*. Kluwer Academic Publishers, Boston, Massachusetts, 1992.

[5] Simon Haykin. *Adaptive Filter Theory*. Prentice-Hall, Englewood Cliffs, New Jersey, 1986.

[6] David J. Heeger. Optical flow from spatiotemporal filters. In *Proc. Int. Conf. on Computer Vision, ICCV-87*, pages 181–190, 1987.

[7] David J. Heeger and Allan Jepson. Simple method for computing 3d motion and depth. In *3rd International Conference on Computer Vision*, pages 96–100, Osaka,Japan, December 4–7 1990.

[8] David J. Heeger and Allan D. Jepson. Subspace methods for recovery of rigid motion I: Algorithm and implementation. *International Journal of Computer Vision*, 7(2):95–117, 1992.

[9] Allan D. Jepson and David J. Heeger. Subspace methods for recovering rigid motion, part II: Theory. Research in Biological and Computational Vision RBCV-TR-90-36, University of Toronto, November 1990.

[10] Allan D. Jepson and David J. Heeger. A fast subspace algorithm for recovering rigid motion. In *Proceedings of the IEEE Workshop on Visual Motion*, pages 124–131, Princeton, New Jersey, October 1991.

[11] Allan D. Jepson and David J. Heeger. Linear subspace methods for recovering translational direction. In L. Harris and M. Jenkin, editors, *Spatial Vision in Humans and Robots*. Cambridge University Press, 1993. See also: Research in Biological and Computational Vision, Department of Computer Science, University of Toronto, RBCV-TR-90-40, Apr. 1992.

[12] Kenichi Kanatani. *Geometric computation for machine vision*. Clarendon Press, Oxford, England, 1993.

[13] Kenichi Kanatani. Renormalization for unbiased estimation. In *Proceedings of the 4th International Conference on Computer Vision*, pages 599–606, Berlin, Germany, May 11–14 1993.

[14] H. C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. *Proceedings of the Royal Society of London B*, 208:385–397, 1980.

[15] W. James MacLean. *Recovery of Egomotion and Segmentation of Independent Object Motion Using the EM Algorithm*. PhD thesis, University of Toronto, 1996.

[16] J. H. Rieger and D. T. Lawton. Processing differential image motion. *J Opt Soc Am A*, 2(2):354–359, 1985.

[17] D. Sinclair. Motion segmentation and local structure. In *Proceedings of the 4th International Conference on Computer Vision*, pages 366–373, Berlin, Germany, May 11–14 1993.

[18] Carlo Tomasi and Takeo Kanade. Factoring image sequences into shape and motion. In *Proceedings of the IEEE Workshop on Visual Motion*, pages 21–28, Princeton, New Jersey, October 1991.

[19] Juyang Weng, Thomas S. Huang, and Narendra Ahuja. Motion and structure from two perspective views: Algorithms, error analysis, and error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(5):451–476, 1989.