Analytical Results on Error Sensitivity of Motion Estimation from Two Views

Konstantinos Daniilidis, Hans-Hellmut Nagel
Fraunhofer Institut für Informations- und Datenverarbeitung
Fraunhoferstr. 1, D-7500 Karlsruhe 1, FRG

Abstract

Fundamental instabilities have been observed in the performance of the majority of the algorithms for three dimensional motion estimation from two views. Many geometric and intuitive interpretations have been offered to explain the error sensitivity of the estimated parameters. In this paper, we address the importance of the form of the error norm to be minimized with respect to the motion parameters. We describe the error norms used by the existing algorithms in a unifying notation and give a geometric interpretation of them. We then explicitly prove that the minimization of the objective function leading to an eigenvector solution suffers from a crucial instability. The analyticity of our results allows us to examine the error sensitivity in terms of the translation direction, the viewing angle and the distance of the moving object from the camera. We propose a norm possessing a reasonable geometric interpretation in the image plane and we show by analytical means that a simplification of this norm leading to a closed form solution has undesirable properties.

1 Introduction

The problem of estimating relative motions between the camera and objects in the scene space from monocular image sequences has been an intensive research area in machine vision during the last years (see the recent survey by [Aggarwal & Nandhakumar 88]). Three different groups of approaches have been developed for the recovery of three dimensional motion from monocular image sequences.

The *discrete* approach is based on the extraction and tracking of features in the image corresponding to three dimensional features in the scene space. The input to the motion algorithm is a set of interframe feature correspondences. The estimated parameters are the rotation and the direction of translation of a three dimensional feature configuration relative to the camera. In the case of points as features, the discrete case is equivalent to the problem of relative orientation in photogrammetry. When the time interval between successive frames is short, the point correspondences yield displacement vectors which may be considered as approximations of the displacement rates used in the continuous approach.

The *continuous* approach is based on an evaluation of the motion or displacement rate field representing the two dimensional velocity of the projections of the three dimensional points. Under certain conditions ([Nagel 89,Girosi et al. 89]) this motion field is equivalent to the optical flow field which describes the apparent instantaneous shift of gray value structures in the image plane. The output of the continuous approaches is the relative angular velocity and the direction of the relative translational velocity of a component of the scene with respect to the camera. A unifying description of approaches for optical flow estimation was presented by [Nagel 87] who also pointed out the relation between the continuous formulation and the discrete approach in case of simple features such as gray value extrema, corners, or edge elements. Underlying the majority of the approaches is the assumption of rigidity of the moving scene component or of the stationary environment in case of a moving camera.

Although the problem of motion estimation from point correspondences or displacement rates from two views was mathematically formulated long ago [Longuet-Higgins 81] there is still no algorithm which has a robust behaviour regardless of motion and scene configuration. In this paper, we study the relation between error sensitivity, the type of motion and the geometry of the problem. It is of major interest to recognize whether the instability depends on the formulation of the problem or if it is inherent in the problem. Our main contribution consists in the analyticity of our results which are consistent with what has already been stated intuitively or observed experimentally. The objective of our study is to point out an instability depending on the formulation of the problem. We begin therefore, with a review –based on unified notation– of the error norms proposed by the existing 3D-motion algorithms.

2 Unifying description of the existing error norms

We denote matrices by capital italics and vectors by boldface symbols. We choose a coordinate system $OXYZ$ with the origin at the center of projection (camera pinhole) and the Z axis coinciding with the optical axis of the camera. The unit vector in the direction of the optical axis is denoted by $\mathbf{z}_0$. We assume that the focal length is unity so that the equation of the image plane xy is $Z = 1$. All the geometric quantities are measured in focal length units. The position vectors of points in the scene space are denoted by uppercase letters $\mathbf{X} = (X, Y, Z)^T$ whereas position vectors of points in the image plane are denoted by lowercase letters $\mathbf{x} = (x, y, 1)^T$. They are related to each other by the equations of perspective projection under the assumption of unit focal length. The notation for the motion parameters is as follows: $\mathbf{t} = (t_x, t_y, t_z)^T$ for the translation, $\mathbf{v} = (v_x, v_y, v_z)^T$ for the translational velocity, R for the rotation and $\omega = (\omega_x, \omega_y, \omega_z)^T$ for the angular velocity. We denote by $\mathbf{x1}, \mathbf{x2}$ and $\mathbf{X1}, \mathbf{X2}$ the points in the image plane and in the scene space at the time instants t_1 and t_2, respectively. The 3D velocity of a point and the displacement rate of an image point are denoted by $\dot{\mathbf{X}}$ and $\dot{\mathbf{x}}$, respectively.

In the discrete case, the rigid motion of an object can be described as a rotation around an axis passing through the origin of the fixed camera coordinate system, followed by a translation. The motion equation of a point on this object reads $\mathbf{X2} = R\mathbf{X1} + \mathbf{t}$. The same equation describes the case of an object being stationary and a camera undergoing a rotation R^T about an axis through the origin followed by a translation $-\mathbf{t}$. We can derive directly a geometric relation between the measurements $\mathbf{x1}$ and $\mathbf{x2}$ and the unknown motion parameters by observing that the two line-of-sight vectors must lie in the same plane with the translation vector $\mathbf{t}$. This coplanarity condition can be written

$$\mathbf{x2}^T(\mathbf{t} \times R\mathbf{x1}) = 0 \tag{1}$$

[Longuet-Higgins 81] and [Tsai & Huang 84] first proposed linear solution methods based on (1) and using *essential parameters* and [Horn 88] proposed the direct minimization of the sum of the squares $\|\mathbf{x2}^T(\mathbf{t} \times R\mathbf{x1})\|$ with respect to the parameters of translation and rotation. All these approaches use the same error term in their minimization function, leading to highly sensitive estimates for the motion parameters as we shall see in the next section.

We continue with the problem formulation in the continuous case. Using the notation given at the beginning of this section, the three dimensional instantaneous velocity of a point on a moving object in scene space is given by

$$\dot{\mathbf{X}} = \mathbf{v} + \omega \times \mathbf{X} \tag{2}$$

After differentiating the equation of the perspective projection $\mathbf{x} = \mathbf{X}/(\mathbf{z}_0{}^T\mathbf{X})$ and making use of (2), we obtain the following expression for the displacement rate:

$$\dot{\mathbf{x}} = \frac{1}{\mathbf{z}_0{}^T\mathbf{X}}\mathbf{z}_0 \times (\mathbf{v} \times \mathbf{x}) \;+\; \mathbf{z}_0 \times (\mathbf{x} \times (\mathbf{x} \times \omega)) \tag{3}$$

We restrict our description and instability considerations to approaches that do not make use of the spatial or temporal derivatives of the image velocities. The displacement rate given in equ. (3) consists of a translational component dependent on the 3D scene structure and the translation, and a rotational component depending only on the angular velocity. If we rewrite the latter as $(\mathbf{z}_0{}^t(\mathbf{x} \times \omega))\,\mathbf{x} - \mathbf{x} \times \omega$ and take the scalar product of (3) with $(\mathbf{v} \times \mathbf{x})$ we obtain the following relation which is free from an explicit dependence on 3D scene structure:

$$(\mathbf{v} \times \mathbf{x})^T\,(\dot{\mathbf{x}} - \omega \times \mathbf{x}) = 0 \tag{4}$$

The same relation is valid when the model of spherical projection $\mathbf{x} = \mathbf{X}/\|\mathbf{X}\|$ is used. In this case [Maybank 86] proposed as an error norm the following residual linear in ω:

$$\epsilon^2(\mathbf{v}) = \min_{\omega} \sum_{i=1}^{m} \left\{(\mathbf{x}_i \times (\mathbf{v} \times \mathbf{x}_i))^T \omega - (\mathbf{v} \times \mathbf{x}_i)^T \dot{\mathbf{x}}_i\right\}^2 \tag{5}$$

A search on the unit hemisphere for the direction of translation minimizing $\epsilon(\mathbf{v})$ was employed to complete the motion estimation. [Zhuang et al. 88] used (4) to formulate a minimization problem based on *essential parameters* (analogous to the discrete case). [Bruss & Horn 83] started by minimizing the discrepancies between the measured and the expected displacement rates with respect to the 3D velocities and the depths:

$$\iint_D \left\{\dot{\mathbf{x}} - \frac{1}{\mathbf{z}_0{}^T\mathbf{X}}\mathbf{z}_0 \times (\mathbf{v} \times \mathbf{x}) - \mathbf{z}_0 \times (\mathbf{x} \times (\mathbf{x} \times \omega))\right\}^2 d\mathbf{x} \Rightarrow \min \tag{6}$$

After eliminating the depths they obtained the objective function

$$\iint_D \left\{\frac{(\mathbf{v} \times \mathbf{x})^T(\dot{\mathbf{x}} - \omega \times \mathbf{x})}{\|\mathbf{z}_0 \times (\mathbf{v} \times \mathbf{x})\|}\right\}^2 d\mathbf{x} \tag{7}$$

which is also proposed by [Scott 88] and is equivalent to the norm we propose. In order to derive a closed form solution, [Bruss & Horn 83] proposed another error norm by weighting the error terms of (7) with $\|\mathbf{z}_0 \times (\mathbf{v} \times \mathbf{x})\|$ yielding

$$\iint_D \left\{(\mathbf{v} \times \mathbf{x})^T(\dot{\mathbf{x}} - \omega \times \mathbf{x})\right\}^2 d\mathbf{x} \tag{8}$$

which can be derived directly from (4). This norm simplifies the minimization but causes a bias in the solution as we demonstrate in the next section.

Although the performance of most algorithms for motion recovery have been tested with synthetic or real data, the dependence of their instability on the form of the objective function, the kind of motion, and the geometry of the problem has been seldom investigated thoroughly. Likewise, very few error analyses lead to explicit formulations of the error sensitivity in terms of the input error, the motion and the structure of the scene. [Adiv 89] pointed out inherent ambiguities in determining the motion of a planar patch. In particular he discovered that, when the field of view is small, it is impossible to distinguish a pure translational motion parallel to the image plane ($\mathbf{v} = (p_1, p_2, 0)^T$) from a pure rotational motion about an axis parallel to the image plane ($\omega = (-p_2/d, p_1/d, 0)^T$). Furthermore, he analytically demonstrated the importance of depth variation and of the ratio of the translation magnitude to the distance of the object from the camera. The instability caused by these factors can be illustrated as a flattening of the error function in the neighborhood of the global minimum. Explicit error analysis has been carried out by [Maybank 86] concerning the error norm (5). He found out that the minima of the residual (5) should lie in the neighborhood of a particular line on the unit sphere and that the ratio v_x/v_y as well as the component of the angular velocity parallel to $(v_x, v_y, 0)^T$ could be reliably estimated in the presence of noise. The influence on the error sensitivity of the translation direction has been emphasized by [Weng et al. 89b], [Horn & Weldon 88] by geometric arguments. The importance of the translation direction has been shown experimentally by [Mitiche et al. 87], [Adiv 89] and [Weng et al. 89b]. [Weng et al. 89b] carried out an analysis of the errors in the estimated eigenvectors in order to compare it with the actual error. However, they did not show any explicit dependence of the error on the motion and structure parameters. [Weng et al. 89a] recognized the importance of a correct error norm and introduced a second step containing a nonlinear minimization of the discrepancies between observed and expected value in a maximum likelihood scheme. They have been able to show that this algorithm achieved the Cramer-Rao lower bounds on the error covariance. They also proposed as an alternative the minimization with respect to the parameters of rotation and translation of the following objective function ([Weng et al. 89a]):

$$\sum_{i=1}^{m} \frac{\left(\mathbf{x2}_i^T(\mathbf{t} \times R\mathbf{x1}_i)\right)^2}{\sigma^2\left(\|\mathbf{z}_0 \times R^T(\mathbf{t} \times \mathbf{x2}_i)\|^2 + \|\mathbf{z}_0 \times (\mathbf{t} \times R\mathbf{x1}_i)\|^2\right)} \tag{9}$$

where σ^2 is the noise variance of the image coordinates. The denominator is the variance of the numerator where the first square error norm in the denominator reflects the uncertainty of $\mathbf{x1}$ and the second one the uncertainty of $\mathbf{x2}$.

3 Error sensitivity concerning the form of the error norm

In this section we show that the sensitivity to the direction of translation and the field of view is not an inherent instability in motion recovery from two views, but depends on the error norm used in the minimization. We explicitly prove that the form of the objective function

$$\iint_D \left\{(\mathbf{v} \times \mathbf{x})^T(\dot{\mathbf{x}} - \omega \times \mathbf{x})\right\}^2 d\mathbf{x} \tag{10}$$

extracted from condition (4) is the reason for the instability in the case of large deviation of the translation direction from the position vector to the object. In (10) we assume that the field of displacement rates $\dot{\mathbf{x}}$ is dense so that we can integrate over an area D with size proportional to the field of view in case of a moving camera or to the size of the projection of the moving object if the camera is stationary.

We are interested in the error of the translation direction, so we restrict ourselves to the case of pure translational motion. Our arguments are also valid in the case of general motion with known rotation or in the case of general motion when the used algorithm estimates translation and rotation sequentially as in [Spetsakis & Aloimonos 88]. When another general motion algorithm is used our arguments build only cues for the experimentally observed instability in the translation direction.

We begin with a geometric justification of our conjecture. Referring to Fig. 1 we can express the deviation of the displacement rate $\dot{\mathbf{x}}$ from the line joining the focus of expansion and the point $\mathbf{x}$ as

$$d' = \frac{|\dot{\mathbf{x}}^T(\mathbf{v} \times \mathbf{x})|}{\|\mathbf{z}_0 \times (\mathbf{v} \times \mathbf{x})\|}$$

We see that what is really minimized when (10) is used is a multiple of the real error distance d'. In fact, the real distance is weighted by the distance of the focus of expansion from the point $\mathbf{x}$. In case of a small field of view, we expect the estimated translation to be biased to the direction of the line of sight to the object. The same geometric justification is valid in the discrete translational case if we assume that only the points $\mathbf{x2}_i$ in the second image are corrupted by noise. Then the distance of $\mathbf{x2}$ from the epipolar line in the second image plane is given by

$$d'' = \frac{|\mathbf{x2}^T(\mathbf{t} \times \mathbf{x1})|}{\|\mathbf{z}_0 \times (\mathbf{t} \times \mathbf{x1})\|}$$

which agrees with the norm used by [Weng et al. 89a] in (9) under the assumption of noise-free $\mathbf{x1}$.

The minimization of (10) in the translational case can be formulated as an eigenvalue problem ([Bruss & Horn 83, Spetsakis & Aloimonos 88] and [Zacharias et al. 85]). The form to be minimized is

$$\iint_D \left\{(\mathbf{v} \times \mathbf{x})^T\dot{\mathbf{x}}\right\}^2 d\mathbf{x} = \mathbf{v}^T\left\{\iint_D (\mathbf{x} \times \dot{\mathbf{x}})^T(\mathbf{x} \times \dot{\mathbf{x}})\, d\mathbf{x}\right\}\mathbf{v} \Rightarrow \min_{\|\mathbf{v}\|=1} \tag{11}$$

The solution for $\mathbf{v}$ is the eigenvector of the matrix

$$A = \iint_D (\mathbf{x} \times \dot{\mathbf{x}})^T(\mathbf{x} \times \dot{\mathbf{x}})\, d\mathbf{x}$$

corresponding to the smallest eigenvalue. The instability of the problem can be expressed in terms of the relative positions between the smallest eigenvalues. In the noise-free case the smallest eigenvalue must be equal to zero. In the presence of noise, the eigenvectors of A can be interpreted as the directions of the axes of the ellipsoid $\mathbf{v}^T A \mathbf{v} = \lambda_{min}$ (see [Bruss & Horn 83]).

The lengths of the axes are 1, $\sqrt{\lambda_1/\lambda_2}$, and $\sqrt{\lambda_1/\lambda_3}$ if we assume without loss of generality that λ_1 is the smallest eigenvalue. Hence, the ellipsoid is circumscribed by a unit sphere with two tangential points on the axis corresponding to the solution eigenvector. The sensitivity to error in the direction of the eigenvector (and hence of the translation) grows when the two smallest eigenvalues come close to each other and the ellipsoid is near to become tangential to the unit sphere in two positions. In this extremely unstable case, a small deformation of the data matrix A can cause a change of 90 degrees in the eigenvector corresponding to the smallest eigenvalue. The instability due to purely isolated eigenvalues has already been investigated in numerical analysis [Wilkinson 65,Golub & van Loan 83]. The perturbation of eigenvectors with respect to the relative position of the eigenvalues is given by [Peters & Wilkinson 74]. We reformulate their result for the case of a symmetric matrix A perturbed by E:

$$\mathbf{x}'_k \approx \mathbf{x}_k + \sum_{j\neq k} \frac{\mathbf{x}_j^T E \mathbf{x}_k}{(\lambda_j - \lambda_k)} \mathbf{x}_j \tag{12}$$

where $\mathbf{x}'_k$ denotes the perturbed eigenvector. From (12) follows that the smaller the difference between the eigenvalues is, the greater is the perturbation in the eigenvector. This result is of general interest since it can be used for the estimation of the error covariance of the solution by all eigenvector problems.

We have been able to prove that this extreme situation of a 90 degrees error in the estimated translation direction can happen in our problem of motion recovery from two views as follows:

We model the projection of the moving object as a rectangle with sides of length α and β in the image plane with its center on the Z axis. Let $\mathbf{v} = (v_x, v_y, v_z)^T$ be the translation of a frontal plane lying at a distance $1/d$ from the origin. By introducing an additional noise term $(\xi, \eta, 0)^T$, the measured displacement rates can be written

$$\dot{\mathbf{x}} = \begin{pmatrix} d(v_x - xv_z) + \xi \\ d(v_y - yv_z) + \eta \\ 0 \end{pmatrix} \tag{13}$$

Using this model we integrate over D and we obtain the elements of A

$$\begin{aligned} a_{11} &= \alpha\beta \left(d^2v_y^2 + d^2v_z^2\frac{\beta^2}{12} + \eta^2 + 2\eta d v_y \right) \\ a_{12} &= \alpha\beta \left(-d^2v_xv_y - dv_x\eta - dv_y\xi - \eta\xi \right) \\ a_{22} &= \alpha\beta \left(d^2v_x^2 + d^2v_z^2\frac{\alpha^2}{12} + \xi^2 + 2\xi d v_x \right) \\ a_{13} &= -\frac{\alpha\beta^3}{12}(d^2v_xv_z + dv_z\xi) \\ a_{23} &= -\frac{\alpha^3\beta}{12}(d^2v_yv_z + dv_z\eta) \\ a_{33} &= \alpha\beta \left(\frac{\alpha^2}{12}(dv_y + \eta)^2 + \frac{\beta^2}{12}(dv_x + \xi)^2 \right) \end{aligned} \tag{14}$$

Suppose that the additional error term $(\xi, \eta, 0)^T$ is an unbiased random variable with second moments $E[\xi^2] = E[\eta^2] = \sigma^2$ and $E[\xi\eta] = 0$. Hence, the matrix A is a random variable with mean

$$A' = E[A] = \alpha\beta \begin{pmatrix} d^2v_y^2 + d^2v_z^2\frac{\beta^2}{12} + \sigma^2 & -d^2v_xv_y & -d^2v_xv_z\frac{\beta^2}{12} \\ -d^2v_xv_y & d^2v_x^2 + d^2v_z^2\frac{\alpha^2}{12} + \sigma^2 & -d^2\frac{\alpha^2}{12}v_yv_z \\ -d^2v_xv_z\frac{\beta^2}{12} & -d^2\frac{\alpha^2}{12}v_yv_z & \frac{\sigma^2}{12}(\alpha^2+\beta^2) + \frac{d^2}{12}(\alpha^2v_y^2 + \beta^2v_x^2) \end{pmatrix} \tag{15}$$

If $\sigma^2 = 0$ then the matrix A' has rank equal to two except in the case when the area D degenerates to a line. In the presence of noise, the matrix has full rank and the smallest eigenvalue is different from zero. The question is whether the two smallest eigenvalues of this matrix are well isolated. Since all the quantities are measured in focal length units, the noise variance is much smaller than unity and the largest value for the sides of the rectangle α and β is two units which happens only when the rectangle covers the whole image and the field of view is 90 degrees.

In the case $v_x = v_y = 0$ the matrix A' takes the diagonal form

$$A' = \alpha\beta \begin{pmatrix} d^2 v_z^2 \frac{\beta^2}{12} + \sigma^2 & 0 & 0 \\ 0 & d^2 v_z^2 \frac{\alpha^2}{12} + \sigma^2 & 0 \\ 0 & 0 & \frac{\sigma^2}{12}(\alpha^2 + \beta^2) \end{pmatrix} \tag{16}$$

and its diagonal elements are identical with the eigenvalues. We take the differences between them

$$\begin{aligned} \lambda_1 - \lambda_3 &= \frac{\alpha\beta^3}{12} d^2 v_z^2 + \frac{\alpha\beta}{12}\sigma^2(12 - \alpha^2 - \beta^2) > 0 \\ \lambda_2 - \lambda_3 &= \frac{\alpha^3\beta}{12} d^2 v_z^2 + \frac{\alpha\beta}{12}\sigma^2(12 - \alpha^2 - \beta^2) > 0 \end{aligned} \tag{17}$$

and observe that the smallest eigenvalue is λ_3 and the corresponding eigenvector is $(0,0,1)^T$ as expected. The differences are significantly greater than zero, the eigenvalues are well isolated and thus the solution is robust.

In the case of a translation parallel to the image plane ($v_z = 0$) we obtain

$$A' = \alpha\beta \begin{pmatrix} d^2 v_y^2 + \sigma^2 & -d^2 v_x v_y & 0 \\ -d^2 v_x v_y & d^2 v_x^2 + \sigma^2 & 0 \\ 0 & 0 & \frac{\sigma^2}{12}(\alpha^2 + \beta^2) + \frac{d^2}{12}(\alpha^2 v_y^2 + \beta^2 v_x^2) \end{pmatrix} \tag{18}$$

The eigenvalues $\lambda_1 < \lambda_2$ are roots of the equation

$$\lambda^2 - \alpha\beta(d^2 v_y^2 + d^2 v_x^2 + 2\sigma^2)\lambda + \alpha^2\beta^2\sigma^2(d^2 v_y^2 + d^2 v_x^2 + \sigma^2) = 0 \tag{19}$$

and read as follows:

$$\lambda_1 = \sigma^2\alpha\beta \qquad \lambda_2 = \alpha\beta(d^2 v_x^2 + d^2 v_y^2 + \sigma^2) \tag{20}$$

We inspect the differences again:

$$\begin{aligned} \lambda_2 - \lambda_1 &= \alpha\beta d^2(v_y^2 + v_x^2) > 0 \\ \lambda_3 - \lambda_1 &= \frac{\alpha\beta}{12}\left(d^2(\alpha^2 v_y^2 + \beta^2 v_x^2) + \sigma^2(\alpha^2 + \beta^2 - 12)\right) \end{aligned} \tag{21}$$

and set the second difference equal to zero in order to obtain the condition for the extreme instability:

$$\sigma^2 = \frac{d^2(\alpha^2 v_y^2 + \beta^2 v_x^2)}{12 - (\alpha^2 + \beta^2)} \tag{22}$$

We show by example that this situation is a realistic one: Let the ratio of the translation magnitude to the distance of the object from the camera be 1/10 and the sides of the rectangle be one tenth of the focal length. It turns out that the noise variance should be about $\sigma^2 = 10^{-4}/12$ — for instance a uniform distribution in the interval $(-0.005, 0.005)$ measured in focal length units. In this case, λ_3 becomes the smallest eigenvalue and we obtain $(0,0,1)^T$ as solution for the translation direction which is wrong by 90 degrees. Thus we have proved that

> *an appropriate combination of noise level, viewing angle, and ratio of the translation magnitude to the distance of the object from the camera can cause an error of 90 degrees in the translation direction if the object is moving parallel to the image plane.*

This error is not to be confused with the two-fold ambiguity in recovering the motion and the normal of the planar patch. This ambiguity allows, as a second interpretation, the interchanging of the translation direction with the normal to the frontal plane accompanied by a rotational motion. This ambiguity does not concern our error analysis since the considered objective function (11) possesses a unique minimum in the unit hemisphere due to the assumption of pure translation.

We now examine what happens between the stable and the extremely unstable situation. We restrict the translation direction to the XZ plane. We set v_y equal to zero in (15) and obtain

$$A' = E[A] = \alpha\beta \begin{pmatrix} d^2v_z^2\frac{\beta^2}{12} + \sigma^2 & 0 & -d^2v_xv_z\frac{\beta^2}{12} \\ 0 & d^2v_x^2 + d^2v_z^2\frac{\alpha^2}{12} + \sigma^2 & 0 \\ -d^2v_xv_z\frac{\beta^2}{12} & 0 & \frac{\sigma^2}{12}(\alpha^2+\beta^2) + \frac{d^2}{12}\beta^2v_x^2 \end{pmatrix} \tag{23}$$

Let λ_2 be the eigenvalue equal to element a'_{22} of the matrix A'. We are interested in the eigenvalues λ_1 and λ_3 which are roots of the characteristic equation

$$\lambda^2 - (a'_{11} + a'_{33})\lambda + a'_{11}a'_{33} - a'^2_{13} = 0 \tag{24}$$

and give rise to eigenvectors lying in the XZ plane. In order to investigate whether they are well isolated, we form the difference between them and after some manipulation we obtain

$$|\lambda_1 - \lambda_3| = \frac{\alpha\beta}{12}\left((d\beta)^4(v_z^2+v_x^2)^2 + \sigma^4(12-\alpha^2-\beta^2)^2 + 2d^2\beta^2\sigma^2(12-\alpha^2-\beta^2)(v_z^2-v_x^2)\right)^{\frac{1}{2}}$$

Let ϕ be the angle between the Z axis and the direction of translation, $G = d^2\beta^2\|\mathbf{v}\|^2$ and $F = \sigma^2(12-\alpha^2-\beta^2) > 0$. Then we have

$$|\lambda_1 - \lambda_3| = \frac{\alpha\beta}{12}\left(G^2 + F^2 + 2FG\cos 2\phi\right)^{1/2} \tag{25}$$

We find that the absolute difference of the eigenvalues decreases as ϕ increases from zero ($\mathbf{v}$ parallel to the optical axis) to 90 degrees ($\mathbf{v}$ parallel to the image plane). The smaller this difference, the greater the perturbations in the estimated eigenvector due to (12). Hence, we can state that the stability of the translation estimation is explicitly related to the translation direction as given by (25).

4 Appropriate Error Norms

To remove the instability proved in the last section, we propose the use of the correct error norm as illustrated in Fig. 1. This norm represents the projection of the translational component of the displacement rate in the direction of the normal to the line passing through the focus of expansion and the point:

$$\iint_D \left\{\frac{(\mathbf{v}\times\mathbf{x})^T(\dot{\mathbf{x}} - \omega\times\mathbf{x})}{\|\mathbf{z}_0\times(\mathbf{v}\times\mathbf{x})\|}\right\}^2 d\mathbf{x} \tag{26}$$

If the spherical projection model is used, the error term should be the component orthogonal to the plane defined by the translation vector and the position vector of the point considered. Due to the high nonlinearity we have not been able to prove analytically that (26) can lead to an unbiased result with respect to the translation direction. The same error norm is suggested by [Spetsakis & Aloimonos 88], [Harris 87] and [Scott 88]. Other error norms having a geometric interpretation on the image plane have been discussed in [Toscani & Faugeras 87a], [Weng et al. 89a] and [Horn 88]. The drawback of (26) and all these error norms is that they lead to a highly nonlinear minimization problem that needs a good initial guess to converge.

Only [Spetsakis & Aloimonos 88] proposed a simplification of the problem in order to obtain an eigenvector solution. Exploiting the fact that the instability occurs when the viewing angle is small, they replaced the position vector of every point $\mathbf{x}$ in the denominator with the position vector $\mathbf{c}$ of the centroid of the projection area. In our notation, this yields

$$\iint_D \left\{\frac{(\mathbf{v}\times\mathbf{x})^T(\dot{\mathbf{x}} - \omega\times\mathbf{x})}{\|\mathbf{v}\times\mathbf{x}\|}\right\}^2 d\mathbf{x} \approx \iint_D \left\{\frac{(\mathbf{v}\times\mathbf{x})^T(\dot{\mathbf{x}} - \omega\times\mathbf{x})}{\|\mathbf{v}\times\mathbf{c}\|}\right\}^2 d\mathbf{x} = \frac{\mathbf{v}^TA\mathbf{v}}{\mathbf{v}^TC\mathbf{v}} \tag{27}$$

where $C = \mathbf{c}^T\mathbf{c}I - \mathbf{c}\mathbf{c}^T$ and the change in the denominator compared to equ. (26) is due to the spherical projection used by [Spetsakis & Aloimonos 88]. The last fraction would be a generalized Rayleigh quotient with an eigenvector solution if the matrix C were not singular.

We will point out that the simplification (27) causes the stable case of motion parallel to the viewing direction to the object to become unstable. This is due to the observation that the denominator in (27) becomes zero when the translation is parallel to $\mathbf{c}$ whereas the numerator is different from zero due to the presence of noise. Hence, the objective function does not have a minimum in the expected position. We use the same model as in the previous section to prove our objection to this simplification explicitly. For $\mathbf{c} = (0,0,1)^T$ and the case of perspective projection onto an image plane, (27) becomes

$$\iint_D \left\{ \frac{(\mathbf{v}\times\mathbf{x})^T\dot{\mathbf{x}}}{\|\mathbf{z}_0 \times (\mathbf{v}\times\mathbf{c})\|} \right\}^2 d\mathbf{x} = \frac{\lambda_1 v_x^2 + \lambda_2 v_y^2 + \lambda_3 v_z^2}{v_x^2 + v_y^2} \tag{28}$$

where $\lambda_i (i = 1\ldots3)$ are the diagonal elements of A' in (16). It is evident that the correct solution $(0,0,1)^T$ causes the value of the objective function to become unbounded. Spetsakis and Aloimonos proposed a deformation of the matrix $C_\delta = \mathbf{c}^T\mathbf{c}I - (1-\delta)\mathbf{c}\mathbf{c}^T$ in order to overcome its singularity. In case of motion towards the object, the matrix C_δ becomes

$$C_\delta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \delta \end{pmatrix} \tag{29}$$

and the minimization of $\mathbf{v}^TA\mathbf{v}/\mathbf{v}^TC\mathbf{v}$ is reduced to the estimation of the eigenvector corresponding to the smallest eigenvalue of

$$A'_\delta = \alpha\beta \begin{pmatrix} d^2v_z^2\frac{\beta^2}{12} + \sigma^2 & 0 & 0 \\ 0 & d^2v_z^2\frac{\alpha^2}{12} + \sigma^2 & 0 \\ 0 & 0 & \frac{\sigma^2}{12\delta}(\alpha^2+\beta^2) \end{pmatrix} \tag{30}$$

By building the differences one more time

$$\begin{aligned} \lambda_1 - \lambda_3 &= \frac{\alpha\beta^3}{12}d^2v_z^2 + \frac{\alpha\beta}{12}\sigma^2(12 - \frac{\alpha^2+\beta^2}{\delta}) \\ \lambda_2 - \lambda_3 &= \frac{\alpha^3\beta}{12}d^2v_z^2 + \frac{\alpha\beta}{12}\sigma^2(12 - \frac{\alpha^2+\beta^2}{\delta}) \end{aligned} \tag{31}$$

we observe that the smallest eigenvalue is no longer λ_3 since the deformation parameter δ takes very small values. Consequently, the error norm (26) must be used in its original nonlinear form so that we avoid the appearance of new unstable configurations.

5 Concluding Remarks

We emphasized the influence of the form of the error norm in minimization approaches for the recovery of motion parameters from two views. We presented the error norms used by existing solution methods in a unifying notation and we referred to sensitivity results already extracted by other authors. We proved analytically that the objective function used in the majority of the approaches leads to an error of 90 degrees in the estimated translation direction when the viewing angle is small, the translation magnitude is small relative to the distance of the moving object from the camera and the motion is parallel to the image plane. We formulated the instability problem in terms representing the isolation of the computed smallest eigenvalue from the other eigenvalues. By using a model for the displacement rates arising from an arbitrary translating frontal plane, we expressed the error sensitivity as a function of the translation direction, the field of view and the ratio of the translation magnitude to the depth of the viewed object.

The situation we modeled in order to obtain analytical results is realistic. Suppose that a camera fixed on a vehicle is viewing an object at a distance with only one planar side visible (for example a stationary truck or bus). In case the camera is moving towards the object the motion can be robustly estimated but if the camera is moving in a lateral direction the estimated parameters become unreliable. Of course the navigation would not rely on the measurements from only two views. Two frames offer such unreliable information, particularly when the whole amount of motion is small or the situation is inherently ambiguous, that even the use of a correct error term would not improve considerably the results. However, before we proceed with the formulation of multiple-frame motion algorithms, it is important to understand the reasons for the instabilities in the two-view motion estimation. New results about the intrinsic ambiguities [Faugeras & Maybank 89] may even lead to the interpretation of specific unstable situations as motion and structure configurations lying in the neighborhood of multiply interpretable situations.

Acknowledgements

The financial support of the first author by the German Academic Exchange Service (DAAD) is gratefully acknowledged. We would like to thank Dr. Greg Hager for helpful comments on a draft of this paper.

References

[Adiv 89] G. Adiv, Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field, *IEEE Trans. Pattern Analysis Machine Intelligence* PAMI-11 (1989) 477-489.

[Aggarwal & Nandhakumar 88] J.K. Aggarwal, N.Nandhakumar, On the computation of motion from sequences of images - a review, *Proceedings of the IEEE* 76 (1988) 917-935.

[Bruss & Horn 83] A. Bruss, B.K.P. Horn, Passive navigation, *Computer Vision, Graphics, and Image Processing* 21 (1983) 3-20.

[Faugeras & Maybank 89] O.D. Faugeras, S. Maybank, Motion from point matches: multiplicity of solutions, in *Proc. IEEE Workshop on Visual Motion*, March 20-22, 1989, Irvine, CA, pp. 248-255.

[Girosi et al. 89] F. Girosi, A. Verri, V. Torre, Constraints for the computation of optical flow, in *Proc. IEEE Workshop on Visual Motion*, March 20-22, 1989, Irvine, CA, pp. 116-124.

[Golub & van Loan 83] G.H. Golub, C.F. van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, Maryland, 1983.

[Harris 87] C.G. Harris, Determination of ego-motion from matched points, in *Proc. Alvey Vision Conference*, Sept. 15-17, 1987, Cambridge, UK, pp. 189-192.

[Horn 88] B.K.P. Horn, Relative orientation, in *Proc. DARPA Image Understanding Workshop*, April 6-8, 1988, Cambridge, MA, pp. 826-837.

[Horn & Weldon 88] B.K.P. Horn, Direct methods for recovering motion, *Int. Journal of Computer Vision* 2 (1988) 51-76.

[Longuet-Higgins 81] H.C. Longuet-Higgins, A computer program for reconstructing a scene from two projections, *Nature* 293 (Sept. 1981) 133-135.

[Maybank 86] S.J. Maybank, Algorithm for analyzing optical flow based on the least squares method, *Image and Vision Computing* 4 (1986) 38-42.

[Mitiche et al. 87] A. Mitiche, X. Zhuang, R. Haralick, Interpretation of optical flow by rotation decoupling, in *Proc. IEEE Workshop on Computer Vision*, Nov. 3 - Dec. 2, 1987, Miami Beach, FL, pp. 195-200.

[Nagel 87] H.-H. Nagel, On the estimation of optical flow: Relations between different approaches and some new results, *Artificial Intelligence* 33 (1987) 299-324.

[Nagel 89] H.-H. Nagel, On a constraint equation for the estimation of displacement rates in image sequences, *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-11 (1989) 13-30.

[Peters & Wilkinson 74] G. Peters, J.H. Wilkinson, Accuracy of computed eigensystems and invariant subspaces, in B.K.P. Scaife (Ed.), *Studies in Numerical Analysis*, Academic Press, London and New York 1974.

[Scott 88] G.L. Scott, *Local and global interpretation of moving images*, Pitman Publishing, London, 1988.

[Spetsakis & Aloimonos 88] M.E. Spetsakis, J. Aloimonos, Optimal computing of structure from motion using point correspondences, in *Proc. Int. Conf. on Computer Vision*, Dec. 5-8, 1988, Tampa, FL, pp. 449-453.

[Toscani & Faugeras 87a] G. Toscani, O.D. Faugeras, Structure and motion from two perspective views, in *Proc. IEEE Int. Conf. on Robotics and Automation*, Mar.31 - Apr. 2, 1987, Raleigh, North Carolina, pp. 221-227.

[Tsai & Huang 84] R. Y. Tsai, T. S. Huang, Uniqueness and estimation of 3-D motion parameters of rigid bodies with curved surfaces, *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-6 (1984) 13-27.

[Weng et.al. 89a] J. Weng, T.S. Huang, N. Ahuja, Optimal motion and structure estimation, in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, June 4-8, 1989, San Diego, CA, pp. 144-152.

[Weng et al. 89b] J. Weng, T.S. Huang, N. Ahuja, Motion and structure from two perspective views: algorithms, error analysis, and error estimation, *IEEE Trans. Pattern Anal. Machine Intelligence* PAMI-11 (1989) 451-476.

[Wilkinson 65] J.H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford Science Publications, Oxford, 1965.

[Zacharias et al. 85] G.L. Zacharias, A.K. Caglayan, J.B. Sinacori, A model for visual flow-field cueing and self-motion estimation, *IEEE Trans. Systems, Man, and Cybernetics* SMC-15 (1985) 385-389.

[Zhuang et al. 88] X. Zhuang, T.S. Huang, N. Ahuja, R.M. Haralick, A simplified linear optical flow-motion algorithm, *Computer Vision, Graphics, and Image Processing* 42 (1988) 334-344.

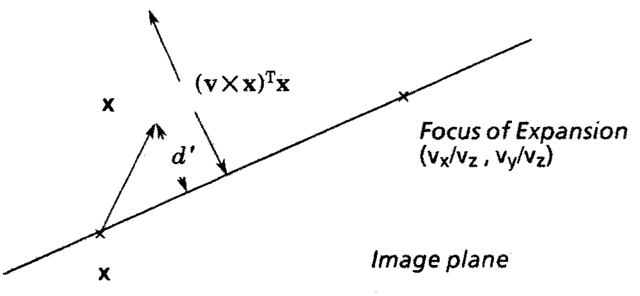

Fig. 1: Geometric interpretation of the error distance d'.