Understanding noise sensitivity in structure from motion

Kostas Daniilidis and Minas E. Spetsakis
Kiel University York University

appeared as Chapter 4, in Y. Aloimonos (Ed.), *Visual Navigation*,
Lawrence Erlbaum Associates, Hillsdale, NJ, 1996, pp. 61-88

Abstract

Solutions to the structure from motion problem have been shown to be very sensitive to measurement noise and the respective motion and geometry configuration. Statistical error analysis has become an invaluable tool in analyzing the sensitivity phenomenon. This paper presents a unifying approach to the problems of statistical bias, correlated noise, choice of error metrics, geometric instabilities and information fusion exploring several assumptions commonly used in motion estimation and reviews several promising techniques for motion estimation. The techniques are based on a small number of principles of statistics and perturbation theory. The analyticity of the approach enables the design of alternatives overcoming the observed instabilities.

1. Introduction

The problem of estimation of structure and motion from moving sequences of images has been attacked by the research community in three stages. The first stage was to answer the question of existence of a solution. This question was answered in the early eighties with algorithms like Longuet-Higgins [1] and independently by Tsai and Huang [2] for the discrete case and Prazdny and Longuet-Higgins [3] and Waxman and Wohn [4] for the differential case. A unique solution was proven to exist and algorithms were developed to find it. The algorithms were practically closed form, simple and easy to implement. Unbeknown to the computer vision community very similar problems had been solved in the context of photogrammetry earlier this century and summarized in standard literature [5].

The second stage was to discover practical algorithms that work under realistic situations were noise is present. The first attempts to solve the problem were frustrated by the high sensitivity to the noise. The opinions in the research community ranged from the pessimistic that the problem is unsolvable, to the optimistic that a few simple heuristics would be sufficient to obtain a general solution. In fact photogrammetrists do have practical algorithms for their needs but these are not directly applicable to most computer vision problems. They depend on extremely high quality cameras, quite often film cameras and they depend on human intervention, to guide the matching process and to take the image sequence. In order for structure from motion to become a general purpose tool, it has to be able to work with general purpose cameras that have very limited resolution, work fast on inexpensive computers and do it with almost no human intervention.

The limited resolution and low quality of the output of ordinary video cameras exaggerates the instabilities in the algorithms for the recovery of the structure and motion. The two interrelated problems here are the the possibility of more than one solution (the ambiguity problem) under certain viewing circumstances (e.g. when viewing a plane) and the noise instabilities. By itself the ambiguity problem is not so interesting because such situations are highly improbable (a perfect plane with no visible features outside the plane in the whole scene is rare). But being near an ambiguous configuration in not improbable and the definition of nearness becomes more inclusive when the noise level increases. The result is increased noise sensitivity especially when the spurious solution is very close to the real one.

The noise sensitivity is usually expressed as a high variance (mean square difference between the ground truth and the estimate) or a bias (mean difference between the ground truth and the estimate). One approach to the problem is to design algorithms to minimize the effects of noise by using a form of constrained minimization [6, 7], use statistical estimation [8, 9, 10], analyze the behavior of the error in a wide variety of configurations [11, 12, 13], analyze the Cramer-Rao Lower Bound (CRLB) of the estimation [14], study the effects of bias and devising tools to reduce it [15] etc. Most of this work has been collected and presented in a recent book by Kanatani [16].

Other approaches include the direct methods [17, 18] that avoid the first step of computing a form of displacement field (flow or correspondence), the subspace methods [19] which provide an elegant solution that involves little more than a set of convolutions and the decoupling methods [20] where the estimation of structure and motion is decoupled in an elegant way using SVD.

Yet another approach involves using features at a level higher than points. Lines were such a feature that produced a great deal of interest. At the beginning non linear algorithms appeared that could not guarantee unique solutions [21] but later with linearized algorithms it was possible to get unique solution almost always [22]. The question whether there exists such linearized algorithm that can use both lines and points was answered positively [23]. When using both points and lines three frames are needed because lines alone provide no constraint for motion with two frames. This method used even with points alone has the advantage of fewer ambiguous configurations. The reason is that an ambiguous configuration with two frames, let us say the first and second frame, will have spurious solutions that are almost always different from the spurious solutions obtained from the first and third frames. Thus the likelihood of a particular configuration being very close to an ambiguous one is much lower than when dealing with two frames. Unfortunately while there exist detailed analyses for the ambiguous configurations for two frames there is very little for three frames.

In most approaches the motion estimation involves two steps. The first step uses as input the image sequence itself and produces a displacement field in the form of discrete point (or feature) correspondences or optical flow field. The flow field has to be accompanied with a representation of the expected error because it is not possible to have reliable dense measurements of the flow over the whole image. Thus a set of discrete correspondences can be considered as a flow field that contains data only at places where the certainty of the flow vectors is high or locally maximal. In most literature the flow is considered small, typically fraction of a pixel and the term discrete displacement field is preferred when the flow is higher than that. The second step consists of the estimation of the motion parameters (rotation and translation) in the form of finite displacements (a vector for the translation and a matrix for the rotation) or infinitesimal displacements (linear and angular velocities). There are algorithms that work either with flow and produce infinitesimal displacements or with discrete matches and produce finite motion. In terms of noise sensitivity and instability, the two general approaches are qualitatively equivalent and they differ mainly in the clarity that different ideas can be presented. In this chapter we are going to use the one most appropriate for each case.

The preferred approach in the study of noise sensitivity is to use analytical rather than numerical tools because of the greater freedom they provide. Simple analytical expressions provide direct insight and more complicated ones can be plotted with various parameters.

The third stage is about how to use the output in conjunction with other perceptory information or cognitive processes to achieve intelligent behavior. As mentioned above the success of the photogrammetric approach is partially due to the intelligent gathering of the data (image sequences) by a skilled operator. An autonomous robotic system has to incorporate techniques that are known as *active vision* [24], to collect data that are more reliable. A great contributor to the mistrust of structure from motion algorithms for real applications comes from the fact that their worst case behavior is very bad. But using some intelligence in the gathering of data, the worst case situations can be avoided. As shown in [24], several ill posed problems become tractable using this paradigm. A related approach is the *purposive vision* [25], where instead of computing the whole structure and motion representation, we can compute only the part that is meaningful to the specific purpose of the system in consideration. This paradigm holds a great promise for the increase of the efficiency of autonomous systems and the reduction of their complexity [26].

The progress of structure from motion cannot be separated from the progress of establishing one or another form of correspondence. The most impressive progress has come from the area of flow computation.

Several novel techniques have been proposed the past several years each one with different advantages and shortcomings. Many of these techniques can be combined with often good results. One of the oldest and simplest techniques by Lucas and Kanade [27] was recently found to outperform most other sophisticated ones [28]. Among the techniques that show great promise when combined and are flexible enough to be combined are the venerable smoothness constraints [29, 30] that, although they do not provide much stability by themselves, they tend to be easy to combine, the hierarchical (or coarse to fine) methods [31] that can help increase the range over which flow can be computed, the statistical techniques proposed by Singh [32] that can relax the Gaussian noise assumption, the affine models of flow that relax the assumption of locally uniform motion [33, 34], the phase based methods that work exceptionally well [35] although they are computationally expensive etc. While flow is usually associated with algorithms with input instantaneous linear and angular velocities, it can be used with any algorithm including the ones that accept as input discrete point correspondences. Using hierarchical techniques, the flow can be computed between images that have much more than one pixel displacement [36].

This chapter is divided into two parts. In the first part we study analytically motion and geometry configurations that yield noise sensitive 3D-information. We present our results and compare them with a state of the art approach [19, 37]. In the second part, we try to overcome some of the instabilities of the first part by paying special attention on the statistical terms of bias and variance. Then, we unify our approach by treating both point and line correspondences in the same way. Last, we analyze the case of recovering the motion parameters from statistically correlated displacement fields.

2. Error analysis of motion estimation

The goal of error analysis is, given the noise level in the measurements, to predict with statistical measures or bounds the error in the motion and depth estimates. These error measures may be computed in closed form or numerically with an algorithm. They express the error in the computed parameters as a function of the true parameters and the measurement noise. Studies like [13, 11, 14, 38, 39] belong to the first category because they contain explicit expressions with respect to the problem parameters. Analyses like those in [10, 40, 41] comprise a second category because they establish bounds or approximations of the estimate error which are computed by sampling the parameter space. Given an expression for the error in the estimate we first study if the error originates in the problem formulation and the algorithm used or if it is inherent in the problem. Since the ultimate goal of error analysis is to find less noise sensitive solutions we are particularly interested in explicit error expressions that will allow either to correct the problem formulation or to actively control the parameter space. The results about the problem formulation –in particular, the form of error metric– described below confirm the fact that the only way to avoid such error is to use the statistical analysis of motion estimation described in the statistical part of this chapter. The way to overcome the problem inherent sensitivities by actively controlling a subset of the parameters is an object of *active* vision we will not describe here.

In this section we will use the instantaneous case because of the easier mathematical treatment due to the linear dependency on the rotational parameters and the presence of measurement noise on the motion field only. In discrete methods the noise can be assumed to corrupt the displacement or each point of the correspondence pair leading to different error measures.

Let an object be moving with translational velocity $\boldsymbol{V} = (V_x, V_y, V_z)^T$ and angular velocity $\boldsymbol{\Omega} = (\Omega_x, \Omega_y, \Omega_z)^T$ relative to the camera. We denote by $\boldsymbol{X}$ the position of a point on the object with respect to the camera coordinate system and by $\boldsymbol{x} = (x, y, 1)^T$ its projection on the image plane $Z = 1$. The direction of the optical axis is given by the unit vector $\hat{\boldsymbol{z}}$. The motion field vector reads [17]

$$\dot{\boldsymbol{x}} = \frac{1}{\hat{\boldsymbol{z}}^T \boldsymbol{X}} \hat{\boldsymbol{z}} \times (\boldsymbol{V} \times \boldsymbol{x}) \; + \; \hat{\boldsymbol{z}} \times (\boldsymbol{x} \times (\boldsymbol{x} \times \boldsymbol{\Omega})). \tag{1}$$

In case of ego-motion of the camera with the above velocities and a stationary environment, the above equation as well as all following equations have to be read with the opposite sign for $\boldsymbol{V}$ and $\boldsymbol{\Omega}$. Due to the linear dependency on the angular velocity $\boldsymbol{\Omega}$ in (1) it is trivial to prove that the error estimate in motion and depth does not depend on the rotational parameters. All the results into which we will delve concern the

case of a small field of view. In particular, we will describe the following facts – proved by other authors or ourselves – that characterize the instability in the 3D-motion estimation from two views:

- A translation can easily be confounded with a rotation in case of small field of view, lateral motion, and not sufficient depth variation [13, 38, 12].
- In case of a small field of view the translation estimate is biased towards the viewing direction if the error metric is not appropriately normalized.
- In case of a small field of view and an irregular surface the cost function to minimize takes its minima along a line in the space of translation directions that goes through the viewing direction and the true translation direction [38, 42]. If the region of interest is fixated this result implies that the azimuth angle of the translation direction can be estimated much more reliable than the polar angle.
- A relation exists between the critical surfaces causing an ambiguity and the critical surfaces causing an instability. This relation provides geometric hints for error sensitivity – among them the translation - rotation confounding mentioned above.

2.1. The translation-rotation confounding

The motion field vector is the sum of two components, a translational that carries the information about the environment and a rotational one (see also eq. (1)). Motions almost parallel to the image plane and in the same direction – like the (V_x, Ω_y) and $(V_y, -\Omega_x)$ pairs – cause a confusion to the observer who cannot disambiguate whether a motion field is induced by a translation or a rotation (see Fig. 1). This confounding becomes dominant if the field of view is small or if there is no depth variation in the environment. However, as already argued by other authors, one may robustly compute the amount of motion represented by the sum $V_x + \Omega_y$ and the difference $V_y - \Omega_x$ since they build the zeroth order terms regarding the motion field as a polynomial with respect to the image coordinates (x, y).

We will first describe the arguments of [38] and then delve into the Cramer-Rao bounds established in [12]. Given the motion field vectors $\dot{\boldsymbol{x}}_{i=1..l}$ in l points we summarize them into an observation vector $\boldsymbol{s}$ of dimension $2l$ and the unknown depths $Z_{i=1..l}$ into the vector of inverse depths $\boldsymbol{Z}$ of dimension l. We may write then the measurement equations (1) for l points as

$$\boldsymbol{s} = C(\boldsymbol{V}) \begin{pmatrix} \boldsymbol{Z} \\ \boldsymbol{\Omega} \end{pmatrix} \qquad \text{with} \qquad C(\boldsymbol{V}) = \begin{pmatrix} A(\boldsymbol{V}) & B \end{pmatrix} \tag{2}$$

where

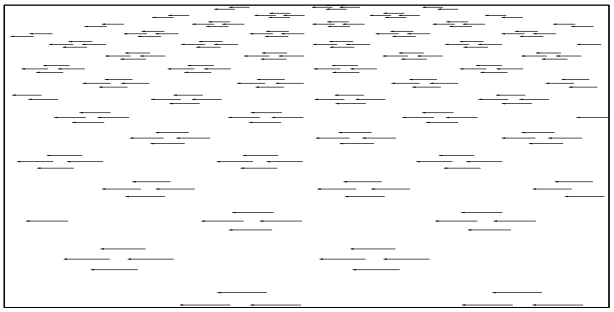
$$A(\boldsymbol{V}) = \begin{pmatrix} V_x - x_1 V_z & \dots & 0 \\ V_y - y_1 V_z & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & V_x - x_l V_z \\ 0 & \dots & V_y - y_l V_z \end{pmatrix} \quad B = \begin{pmatrix} -x_1 y_1 & 1 + x_1^2 & -y_1 \\ -(1 + y_1^2) & x_1 y_1 & x_1 \\ & \vdots & \\ -x_l y_l & 1 + x_l^2 & -y_l \\ -(1 + y_l^2) & x_l y_l & x_l \end{pmatrix}$$

are matrices of dimension $2l \times l$ and $2l \times 3$, respectively. According to [38] the necessary and sufficient condition for an observation vector $\boldsymbol{s}$ to be consistent with a translation $\boldsymbol{V}$ is that

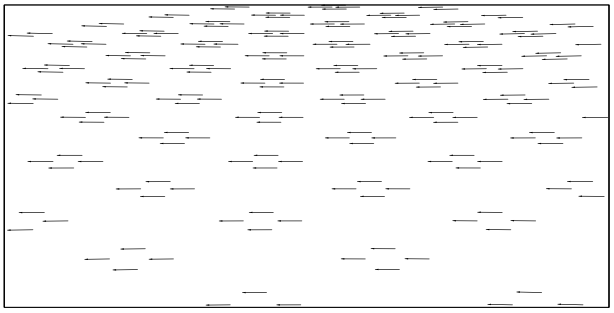
$$\boldsymbol{s} \in \text{range}(C(\boldsymbol{V})). \tag{3}$$

This condition is equivalent to the requirement that the observation vector $\boldsymbol{s}$ is orthogonal to every vector of the orthogonal complement of range$(C(\boldsymbol{V}))$. Since the orthogonal complement of the range of a matrix is equal to the null-space of the transpose of the matrix we have

$$\boldsymbol{s}^T \Psi(\boldsymbol{V}) = 0 \qquad \text{for every} \qquad \Psi(\boldsymbol{V}) \in \text{null}(C(\boldsymbol{V})^T). \tag{4}$$



(a)



(b)

Figure 1: Pure translational (a) and pure rotational (b) motion field induced by V_x-translation and Ω_y-rotation, respectively. A large field of view allows the perception of the depth variation through the change of the flow magnitude in (a). A small field of view around the center contains in both cases (a) and (b) almost identical fields.

The Ψ-vectors are constructed as a linear combination $\sum_{i=1}^{l} c_i(\boldsymbol{V})\Phi_i(\boldsymbol{V})$ of l vectors $\Phi_i(\boldsymbol{V})$ spanning the null-space null$(A(\boldsymbol{V})^T)$ of the submatrix $A(\boldsymbol{V})^T$. The necessary and sufficient condition for $\Psi(\boldsymbol{V})$ to belong to null$(C(\boldsymbol{V})^T)$ is to annihilate the rotational component:

$$B^T \sum_{i=1}^{l} c_i(\boldsymbol{V})\Phi_i(\boldsymbol{V}) = 0 \qquad \text{or} \qquad E(\boldsymbol{V}) \begin{pmatrix} c_1 & \dots c_l \end{pmatrix}^T = 0 \tag{5}$$

with

$$E(\boldsymbol{V}) = B^T \begin{pmatrix} \Phi_1 & \dots & \Phi_l \end{pmatrix}.$$

Hence, the problem is reduced to the computation of the null-space of $E(\boldsymbol{V})$. The degenerate case of deficient rank of $C(\boldsymbol{V})$ occurs if the $\Phi_i(\boldsymbol{V})$ are not independent – this implies that the FOE is on one of the points – or if the matrix $E(\boldsymbol{V})$ has rank less than three. In the latter case, it exists a configuration of depths and rotation $(\boldsymbol{Z}', \boldsymbol{\Omega}')^T$ such that $C(\boldsymbol{V})(\boldsymbol{Z}', \boldsymbol{\Omega}')^T = 0$. This means, that there exists a rotation $\boldsymbol{\Omega}'$ that induces a pure translational motion field:

$$B\boldsymbol{\Omega}' = -A(\boldsymbol{V})\boldsymbol{Z}'.$$

In case of a small field of view it can be proved that the induced depths are planar and the induced translation is perpendicular to the actual one [38].

We next proceed to show the confounding in the space of uncertainty directions as represented by the eigenspace of the Fisher Information matrix. In order to reduce the number of the depth unknowns we sacrifice generality and assume that the perceived surface in motion is planar. Let the plane be given by the equation $\boldsymbol{N}^T\boldsymbol{X} = 1$ where $\boldsymbol{N} = (N_x, N_y, N_z)^T$ has the direction of the normal to the plane and a magnitude equal to the inverse of the distance of the origin to the plane. By dividing by the depth we obtain $1/\hat{\boldsymbol{z}}^T\boldsymbol{X} = \boldsymbol{N}^T\boldsymbol{x}$ which we insert in eq. (1):

$$\dot{\boldsymbol{x}} = (\boldsymbol{N}^T\boldsymbol{x})(\hat{\boldsymbol{z}} \times (\boldsymbol{V} \times \boldsymbol{x})) + \hat{\boldsymbol{z}} \times (\boldsymbol{x} \times (\boldsymbol{x} \times \boldsymbol{\Omega})). \tag{6}$$

Let $\boldsymbol{p}$ be the vector of unknown parameters – in our case motion parameters and the normal, but we will define them later – and $\mathcal{Z}$ be the set of all measurements – in our case all motion field vectors. The *Fisher information matrix* is defined as follows [43]

$$F = E[\frac{\partial \ln p(\mathcal{Z}|\boldsymbol{p})}{\partial \boldsymbol{p}}^T \frac{\partial \ln p(\mathcal{Z}|\boldsymbol{p})}{\partial \boldsymbol{p}}], \tag{7}$$

where $p(\mathcal{Z}|\boldsymbol{p})$ is the conditional probability density function. The uncertainty of an estimator $\hat{\boldsymbol{p}}$ is given by its error covariance $E[(\boldsymbol{p} - \hat{\boldsymbol{p}})(\boldsymbol{p} - \hat{\boldsymbol{p}})^T]$. Following the Cramer-Rao inequality [43], the error covariance of an *unbiased* estimator is bounded from below by the inverse of the Fisher information matrix:

$$E[(\boldsymbol{p} - \hat{\boldsymbol{p}})(\boldsymbol{p} - \hat{\boldsymbol{p}})^T] \geq F^{-1}. \tag{8}$$

The inequality for matrices means that the difference of the lhs minus the rhs is a positive semidefinite matrix. Since the diagonal elements of a positive semidefinite matrix are greater or equal to zero we can directly recover scalar lower bounds for the variances of the unknowns. However, the inverse of the Fisher information matrix provides much richer information about the most and least error sensitive directions in the parameter space. In the optimistic case of an efficient estimator, the uncertainty may be illustrated by the following uncertainty ellipsoid

$$(\boldsymbol{p} - \hat{\boldsymbol{p}})^T F (\boldsymbol{p} - \hat{\boldsymbol{p}}) = c \tag{9}$$

with the estimate being the center of the ellipsoid. The probability that the true value $\boldsymbol{p}$ lies inside the ellipsoid is given by the constant c which geometrically expresses the stretching of the ellipsoid. The directions of the axes of symmetry of the ellipsoid are given by the eigenvectors of F. The lengths of the semi-axes are equal to $\sqrt{(c/\lambda)}$ where λ is the corresponding eigenvalue of F. The direction of the lowest uncertainty is given by the eigenvector corresponding to the largest eigenvalue of F – this is not a contradiction, if one recalls that the error covariance lower bound is equal to the inverse of F. This direction allows us to obtain insight into the question which linear combinations of the unknown parameters (projections of the

parameter vector onto subspaces) can be robustly estimated even if each parameter estimate for itself may have a high uncertainty.

The analytic computation of the Fisher information matrix requires a model of the probability density function of the measurements. We assume a Gaussian distribution for all measured motion field vectors, with zero mean and covariance equal to $\sigma^2 I$. Under the above assumptions it can be proved that the Fisher information matrix reads

$$F = \iint_{I_D} \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{p}}^T \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{p}} \, dx \, dy. \tag{10}$$

We assume a dense motion field over the domain D equal to the area of the projection of the environmental part moving relative to the camera which we call the *effective* field of view. It is equal to the field of view in case of a stationary environment and egomotion of the camera. The measurement function $\boldsymbol{h}(\boldsymbol{p})$ of the motion field (1) is given by

$$\begin{aligned} \dot{\vec{\boldsymbol{x}}} &= D\boldsymbol{q} = \begin{pmatrix} 1 & x & y & 0 & 0 & 0 & x^2 & xy \\ 0 & 0 & 0 & 1 & x & y & xy & y^2 \end{pmatrix} \boldsymbol{q} \\ \text{with} \quad \boldsymbol{q} &= (V_x N_z + \Omega_y,\ V_x N_x - V_z N_z,\ V_x N_y - \Omega_z, \\ &\quad V_y N_z - \Omega_x,\ V_y N_x + \Omega_z,\ V_y N_y - V_z N_z, \\ &\quad \Omega_y - V_z N_x,\ -V_z N_y - \Omega_x)^T. \end{aligned}$$

We used a different symbol $\dot{\vec{\boldsymbol{x}}}$ for the motion field vector in $\mathcal{R}^2$ in contrast to $\dot{\boldsymbol{x}}$ in (1) which belongs to $\mathcal{R}^3$ with the third component equal to zero.

The Jacobian $\partial \boldsymbol{q} / \partial \boldsymbol{p}$ is independent of the image coordinates, hence

$$F = \frac{\partial \boldsymbol{q}}{\partial \boldsymbol{p}}^T \left\{ \iint_{I_D} D^T D \, dx \, dy \right\} \frac{\partial \boldsymbol{q}}{\partial \boldsymbol{p}}. \tag{11}$$

We model the integration domain I_D – i.e. the effective field of view – as a rectangle placed in the image center with side lengths equal to α and β. The integral matrix

$$D_{int} = \iint_{I_D} D^T D \, dx \, dy \tag{12}$$

depends only on the size of the field of view and is a multiple of $\alpha\beta$. Thus, the error covariance is a monotonically decreasing function of the size of the effective field of view. Before we proceed with the computation of the Jacobian $\partial \boldsymbol{q} / \partial \boldsymbol{p}$ we must choose eight independent unknowns among the elements of $\boldsymbol{V}$, $\boldsymbol{\Omega}$ and $\boldsymbol{N}$. We assume that $N_z \neq 0$ which implies that the planar surface is not parallel to the optical axis and we make the following substitutions: $N'_x = N_x/N_z, N'_y = N_y/N_z, V'_x = V_x/N_z, V'_y = V_y/N_z, V'_z = V_z/N_z$. For the sake of simple expressions, we retain the unprimed symbols instead of the primed ones. The vector of independent unknown parameters then reads as follows:

$$\boldsymbol{p} = \begin{pmatrix} V_x & V_y & V_z & \Omega_x & \Omega_y & \Omega_z & N_x & N_y \end{pmatrix}^T.$$

Before we recover the inverse of the Fisher information matrix we compute its determinant:

$$\det(F) = \det^2\left(\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{p}}\right) \det(D_{int})$$

After tedious adding and subtracting the rows of the Jacobian $\partial \boldsymbol{q} / \partial \boldsymbol{p}$ we obtain

$$\det\left(\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{p}}\right) = \|\boldsymbol{N} \times \boldsymbol{V}\|^2. \tag{13}$$

which shows as expected that the case of parallel translation and normal causes a high uncertainty as the degenerate case of the merging of the two solutions to one [44]. We carry out the matrix multiplications in (11) and obtain

$$F = \begin{pmatrix} K & L \\ L^T & M \end{pmatrix} \tag{14}$$

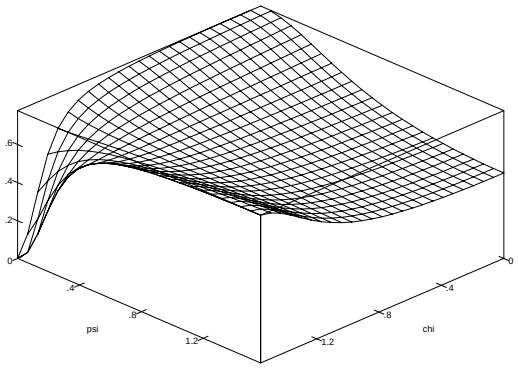


Figure 2: The smallest eigenvalue of S as a function of the angles ψ of translation and χ of the normal with the optical axis for a small field of view: A $= 0.1$

where K, L, M are 6×6, 6×2 and 2×2 matrices, respectively. In order to invert the Fisher information matrix computed in the last section we make use of the formula [45]

$$F^{-1} = \begin{pmatrix} E^{-1} & -E^{-1}LM^{-1} \\ -M^{-1}L^TE^{-1} & M^{-1} + M^{-1}L^TE^{-1}LM^{-1} \end{pmatrix} \tag{15}$$

$$\text{with} \qquad E = (K - LM^{-1}L^T)$$

and F as in (14). The matrix E obtains the following block-diagonal form

$$E = \begin{pmatrix} E_{135} & 0 \\ 0 & E_{246} \end{pmatrix}, \tag{16}$$

if we set $V_y = 0$ and $N_y = 0$. This means that the translational velocity as well as the normal lie on the XZ plane. We introduce the angles ψ and χ between the optical axis and $\boldsymbol{V}$ and $\boldsymbol{N}$, respectively:

$$\begin{array}{rcl} \boldsymbol{V} & = & \begin{pmatrix} V_x & 0 & V_z \end{pmatrix} = \begin{pmatrix} \|\boldsymbol{V}\| \sin\psi & 0 & \|\boldsymbol{V}\| \cos\psi \end{pmatrix} \\ \boldsymbol{N} & = & \begin{pmatrix} N_x & 0 & 1 \end{pmatrix} = \begin{pmatrix} \tan\chi & 0 & 1 \end{pmatrix}. \end{array}$$

The block matrices E_{135} and E_{246} correspond to the unknown triples (V_x, V_z, Ω_y) and $(V_y, \Omega_x, \Omega_z)$, respectively. We are, thus, able to invert the matrix E by inverting the two 3×3 block matrices.

$$E^{-1} = \begin{pmatrix} E_{135}^{-1} & 0 \\ 0 & E_{246}^{-1} \end{pmatrix}. \tag{17}$$

We used the MAPLE symbolic package to compute the inverses E_{135}^{-1} and E_{246}^{-1}. The uncertainty between the two unknown-triples is decoupled. We will study the first triple (V_x, V_z, Ω_y), the study of the second triple can be conducted analogously. We introduce the unit-vector $\boldsymbol{u} = (\cos\phi, 0, \sin\phi)^T$. The quadratic form $\boldsymbol{u}^T E_{135}^{-1} \boldsymbol{u}$ represents the uncertainty in direction ϕ. The uncertainty in the (V_x, Ω_y) space can be illustrated geometrically as the intersection of an ellipsoid in $(V_x, , V_z, \Omega_y)$ space with a plane. Let S be the 2×2 submatrix of E_{135} built by the first and third columns and rows of E_{135}. The bounds of the quadratic form $\boldsymbol{u}^T E_{135}^{-1} \boldsymbol{u}$ are given by the smallest and the largest eigenvalue of S [46]:

$$\lambda_{min}(S) \leq \boldsymbol{u}^T E_{135}^{-1} \boldsymbol{u} \leq \lambda_{max}(S). \tag{18}$$

We are interested in the value of the lowest uncertainty which is proportional to $\lambda_{min}(S)$. The expression for $\lambda_{min}(S)$ computed by MAPLE is very long. We restrict ourselves to plot it as a function of ψ and χ. Fig. 2 shows that the smallest eigenvalue is not affected by the singularity $\chi = \psi$. However, the error variances of V_x and Ω_y become infinitely large. This fact substantiates our methodology in exploiting the

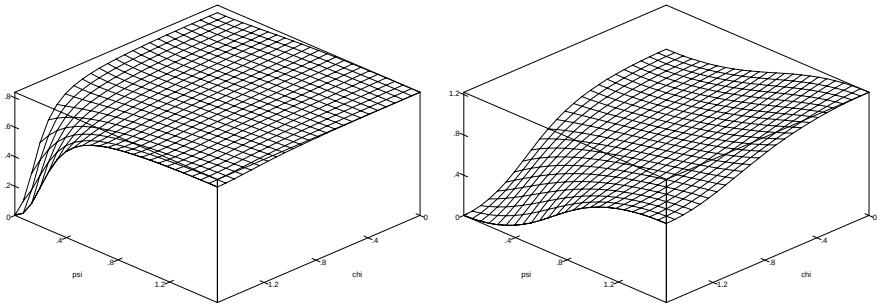


Figure 3: The angle ϕ_{min} of the lowest uncertainty direction as a function of the angles ψ of translation and χ of the normal with the optical axis for two sizes of the field of view: A $= 0.1$ (a) and A $= 1.0$ (b). In case (a) of small field of view this angle is almost everywhere equal $\pi/4$. Thus the direction of lowest uncertainty is $(1, 1)$ and the sum $V_x + \Omega_y$ may be robustly estimated. This effect becomes the weaker the larger the size of the field of view.

entire structure of the lower bound covariance matrix. Fig. 2 shows that the variance in the direction of the lowest uncertainty is an increasing function of the slant χ if the translation is parallel to the image plane ($\psi = \pi/2$) and a decreasing function of the slant χ if the translation is parallel to the optical axis ($\psi = 0$). We next compute the angle ϕ_{min} that gives the direction of lowest uncertainty with the help of MAPLE. Fig. 3 shows that for a small field of view the angle ϕ_{min} takes almost everywhere values close to $\pi/4$. Hence, the direction of lowest uncertainty is $(\cos\pi/4, \sin\pi/4)$ which implies that the sum $V_x + \Omega_y = (1, 1)(V_x, \Omega_y)^T$ can be robustly estimated. Values of ϕ_{min} near zero mean that the most robust direction in (V_x, Ω_y)-space is $(1, 0)$, implying that the estimate for translational velocity V_x is robust. This happens if the plane is parallel to the optical axis (χ near $\pi/2$) and the translation is parallel to the optical axis (ψ near zero) as well. Planes parallel to the optical axis induce a high variation in the magnitudes of the motion field vectors. Translations parallel to the optical axis induce radially expanding motion fields. In both cases the motion field cannot be confused with a motion field induced by a pure rotation about an axis parallel to the image plane. The effect of a dominant direction in (V_x, Ω_y)-space is weaker if the field of view is large (Fig. 3 below). The angle ϕ_{min} may take values greater than $\pi/4$ but is never close to $\pi/2$ which prevents the estimate for Ω_y from having the lowest uncertainty. We have, thus, shown that the directions of the lowest uncertainty in the mixed translational-rotational parameter space correspond to the sum and difference of the components of the velocities causing motion parallel to the image plane. The lower bounds for each component individually are higher and this effect is amplified if the size of the field of view and the slant of the plane become smaller. The only way to attenuate this inherent sensitivity is to keep the amount of lateral translation as small as possible by actively fixating on the Focus of Expansion.

2.2. The bias towards the viewing direction

In this section we are going to prove the bias in the estimated translation direction towards the viewing direction if the field of view is small. This bias was already observed in [47, 48, 49, 37]. We will show here using the arguments in [39, 11] that this bias can be eliminated if the error metric is derived by the statistical analysis in the first part of the chapter. The bias affected error metrics are those derived directly from the epipolar constraint in its discrete form

$$\boldsymbol{x}_2{}^T(\boldsymbol{T} \times R\boldsymbol{x}_1) = 0. \tag{19}$$

or its instantaneous form

$$(\boldsymbol{V} \times \boldsymbol{x})^T(\dot{\boldsymbol{x}} - \boldsymbol{\Omega} \times \boldsymbol{x}) = 0 \tag{20}$$

without considering the noise in $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ or $\dot{\boldsymbol{x}}$, respectively. It should be noted that the above equations are valid in both cases of a planar and a spherical image surface. Furthermore, the bias does not affect

approaches minimizing

$$\iint_{I_D} \left\| \dot{\boldsymbol{x}} - \frac{1}{\hat{\boldsymbol{z}}^T \boldsymbol{X}} \hat{\boldsymbol{z}} \times (\boldsymbol{V} \times \boldsymbol{x}) - \hat{\boldsymbol{z}} \times (\boldsymbol{x} \times (\boldsymbol{x} \times \boldsymbol{\Omega})) \right\|^2 dx\, dy \tag{21}$$

like [19, 50, 51, 52] since the error metric obtained after elimination of depth

$$\iint_{I_D} \left\{ \frac{(\boldsymbol{V} \times \boldsymbol{x})^T (\dot{\boldsymbol{x}} - \boldsymbol{\Omega} \times \boldsymbol{x})}{\|\hat{\boldsymbol{z}} \times (\boldsymbol{V} \times \boldsymbol{x})\|} \right\}^2 dx\, dy, \tag{22}$$

is correctly normalized. The error-metric (22) is equivalent to the error metrics for the discrete case obtained in the first part of the chapter by statistical arguments. On the contrary, it affects the minimization of the error metric

$$\sum_{n=1}^{N} \left(c_n (\boldsymbol{x} \times \dot{\boldsymbol{x}})^T \boldsymbol{V} \right)^2 \tag{23}$$

used in [37] where the coefficient vector $(c_1..c_N)$ is chosen to be orthogonal to all quadratic polynomials in (x, y) so that the quadratic rotational component of the motion field will be annihilated. Since the coefficients c_n are independent of the translation $\boldsymbol{V}$ the above error metric is biased in the same way as the metric $(\boldsymbol{x} \times \dot{\boldsymbol{x}})^T \boldsymbol{V}$.

We proceed with introducing essential parameters into the instantaneous epipolar constraint (20) which may be written as

$$\boldsymbol{V}^T (\boldsymbol{x} \times \dot{\boldsymbol{x}}) + \boldsymbol{x}^T \mathcal{K} \boldsymbol{x} = 0 \tag{24}$$

with $\mathcal{K} = \boldsymbol{V}^T \boldsymbol{\Omega} I - \frac{1}{2}(\boldsymbol{V}\boldsymbol{\Omega}^T + \boldsymbol{\Omega}\boldsymbol{V}^T)$. Ignoring the constraints for the decomposability of $\mathcal{K}$ in $\boldsymbol{V}$ and $\boldsymbol{\Omega}$ temporarily a solution for the minimization of

$$\iint_{I_D} \left(\boldsymbol{V}^T (\boldsymbol{x} \times \dot{\boldsymbol{x}}) + \boldsymbol{x}^T \mathcal{K} \boldsymbol{x} \right)^2 dx\, dy$$

is given by the eigenvector for the smallest eigenvalue of the matrix [39]

$$\mathcal{Y} = \iint_{I_D} (\boldsymbol{x} \times \dot{\boldsymbol{x}})(\boldsymbol{x} \times \dot{\boldsymbol{x}})^T dx\, dy - \mathcal{M}$$

where $\mathcal{M}$ depends only on the image coordinates (x, y). Suppose now that the motion field vectors $\dot{\boldsymbol{x}}$ are corrupted by additive noise

$$\dot{\boldsymbol{x}}' = \dot{\boldsymbol{x}} + \begin{pmatrix} \xi \\ \eta \\ 0 \end{pmatrix}$$

with vanishing expectations and isotropic variances $E[\xi^2] = E[\eta^2] = \sigma^2$ and $E[\xi\eta] = 0$. The expectation of the corrupted matrix $\mathcal{Y}'$ reads then

$$E[\mathcal{Y}'] = \mathcal{Y} + \sigma^2 \begin{pmatrix} \alpha\beta & 0 & 0 \\ 0 & \alpha\beta & 0 \\ 0 & 0 & \alpha\beta\frac{\alpha^2+\beta^2}{12} \end{pmatrix},$$

where α and β are the side lengths of the rectangular domain of integration around the viewing direction assumed to be parallel to the optical axis. The bias in the matrix $\mathcal{Y}$ affects the eigenvector for the smallest eigenvalue and is of order $\mathcal{O}(\sigma^2)$. It increases with decreasing difference between the two smallest eigenvalues according to the well known theorem [46] that the perturbed eigenvector $\boldsymbol{x}'_k$ of a perturbed symmetric matrix $A + E$ reads

$$\boldsymbol{x}'_k \approx \boldsymbol{x}_k + \sum_{j \neq k} \frac{\boldsymbol{x}_j^T E \boldsymbol{x}_k}{(\lambda_j - \lambda_k)} \boldsymbol{x}_j, \tag{25}$$

with λ_i, $\boldsymbol{x}_i$ the eigenvalues and eigenvectors of the unperturbed matrix respectively. Assuming a true motion field arising from a planar surface $N_x X + N_y Y + N_z Z = 1$ it is proved in [11] that

1. The difference between the two smallest eigenvalues is a monotonically decreasing function of the polar angle θ of the true translational direction

$$|\lambda_1' - \lambda_3'| = \left(F^2 + G^2 + 2FG\cos 2\theta\right)^{\frac{1}{2}}, \tag{26}$$

where F and G depend on the noise σ^2, the area of the field of view, the slope of the surface, and the ratio of translation to distance.

2. The upper bound on the bias of the eigenvector expressed as the sine of its rotation is a monotonically increasing function of the polar angle of the translation

$$|\sin\delta\varphi| \leq \frac{\sigma^2\sqrt{\sin^2\theta + (\frac{\alpha^2+\beta^2}{12})^2\cos^2\theta}}{\lambda_3 + \sigma^2\frac{\alpha^2+\beta^2}{12}}. \tag{27}$$

We conclude that the more lateral the true translation and the smaller the field of view, the more severe are the bias effects.

For the elimination of bias it is proposed by Kanatani [39] to minimize the Rayleigh quotient

$$\frac{\boldsymbol{V}^T\mathcal{Y}\boldsymbol{V}}{\boldsymbol{V}^T\mathcal{B}\boldsymbol{V}}$$

where $\mathcal{B}$ is the bias matrix

$$\begin{pmatrix} \alpha\beta & 0 & 0 \\ 0 & \alpha\beta & 0 \\ 0 & 0 & \alpha\beta\frac{\alpha^2+\beta^2}{12} \end{pmatrix}.$$

This can be reduced to a generalized eigenvalue problem $\mathcal{Y}\boldsymbol{V} = \lambda\mathcal{B}\boldsymbol{V}$. This is equivalent to finding the eigenvector for the smallest eigenvalue of the unbiased matrix $\mathcal{Y} - \lambda\mathcal{B}$ where the unknown λ represents the noise level in the motion field. It can be easily proved that the normalizing denominator is equal

$$\boldsymbol{V}^T\mathcal{B}\boldsymbol{V} = \iint_{I_D} \|\hat{\boldsymbol{z}} \times (\boldsymbol{V} \times \boldsymbol{x})\|^2\, dx\, dy.$$

Thus, the bias can be eliminated if we integrate separately the denominator in (22) what is approximately correct if the field of view is very small.

2.3. The important line

In this section we will describe the results obtained in [38, 42] about the form of the error surface around the minimum in the space of translation directions. We should remark at this point that the curvature of the error surface in the neighborhood of the minimum is a measure of the stability of this minimum. The Fisher Information matrix of the last chapter is nothing else than an approximate measure of the flatness of the error surface near the minimum. The error surface used by Jepson and Heeger [38] is the residuum obtained by the pseudoinverse solution of (2):

$$J(\boldsymbol{V}) = \|(I - C(\boldsymbol{V})C^{\dagger}(\boldsymbol{V}))\boldsymbol{s}\|^2. \tag{28}$$

For a sufficiently small field of view and an irregular depth variation the residuum $J(\boldsymbol{V})$ was proved to be proportional to $\sin\theta_o\sin(\phi - \phi_o)$ where

$$\boldsymbol{V} = (\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta)^T$$

and (θ_o, ϕ_o) are the polar and azimuth angle of the actual translation. Thus, the residuum varies only in the direction perpendicular to the line connecting the viewing direction and the true translation. We illustrate this fact in the (θ, ϕ) coordinate system of Fig. 4.

Figure 4: The residuum $J(\boldsymbol{V})$ as a function of the translation direction angles (θ, ϕ). The actual translation is $(1/2, 1/2, 1)$, the depth is a step function, and there is no rotation. The residuum is shown for a small (0.2×0.2) and a large (1×1) field of view. The valley of minima can be clearly seen in case of a small field of view (left).

Figure 5: The residuum $J_m(\boldsymbol{V})$ as a function of the translation direction angles (θ, ϕ). The actual translation is $(1/2, 1/2, 1)$, the depth is a step function, and there is no rotation. The residuum is shown for a small (0.2×0.2) and a large (1×1) field of view. The valley of minima can be clearly seen in case of a small field of view (left). Furthermore, we observe the bias towards the viewing direction ($\theta = 0$) due to the use of the unnormalized epipolar constraint.

A similar relationship was derived be Maybank [42, 53] but the residuum after elimination of depths becomes

$$J_m(\boldsymbol{V}) = \sum_{i=1}^{l} \|(\boldsymbol{V} \times \boldsymbol{x}_i)^T (\dot{\boldsymbol{x}}_i - \boldsymbol{\Omega} \times \boldsymbol{x}_i)\|^2. \tag{29}$$

It was proved that under the same assumptions of small field of view and irregular depth distribution the residuum above is proportional to $\|\boldsymbol{V}^T(\boldsymbol{V}_o \times \boldsymbol{x}_o)\|$ where $\boldsymbol{x}_o$ is the viewing direction and $\boldsymbol{V}_o$ is the true translation. It is trivial to show that the triple product has as a factor $\sin(\phi - \phi_o)$. The same valley of minima can be observed in Fig. 5 with a small but expected surprise: Due to the use of the epipolar constraint as an error metric the locations of minima are biased towards the viewing direction $\theta = 0$.

2.4. Critical surfaces and instability

In this section we switch over again to the discrete case of given point correspondences $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ in order to study the relation between ambiguity and error sensitivity in the structure from motion problem. It has been repeatedly shown in the past that certain configurations of scene points and optical centers – called critical surfaces – cause the existence of two or three solutions given more than five point correspondences (see [53] and the references therein). In the context of error sensitivity we are interested here only in solutions that

differ slightly from each other. Given a solution for translation and rotation we study the geometries for which a first order perturbation of the solution yields the same point correspondences. Horn [54] showed that the epipolar constraint remains unaffected by first order deformations of the motion parameters if the points lie on a quadric with certain properties. The relation of these instability-critical surfaces to the ambiguity-critical surfaces has been an open problem in visual motion[a]. Such a relationship was first established in pure geometric terms by Hofmann [56]. Following [57] we show here that an instability-critical surface arises when the two optical centers take special positions on the ambiguity-critical surface.

Let $(R, \boldsymbol{a})$ and $(S, \boldsymbol{b})$ be two rotation-translation pairs that yield the same point correspondences $(\boldsymbol{x}_1, \boldsymbol{x}_2)$. The pair of the two ambiguity surfaces associated with each of the solutions reads [53]

$$(U\boldsymbol{X} \times \boldsymbol{X})^T \boldsymbol{b} = (U\boldsymbol{a} \times \boldsymbol{X})^T \boldsymbol{b} \quad (30)$$

$$(U^T\boldsymbol{X} \times \boldsymbol{X})^T \boldsymbol{a} = (U^T\boldsymbol{b} \times \boldsymbol{X})^T \boldsymbol{a}, \quad (31)$$

where $U = SR^T$ is the rotation difference and $\boldsymbol{X}$ the scene point with respect to the first coordinate system. The surfaces are quadrics of the form

$$\boldsymbol{X}^T M \boldsymbol{X} + \boldsymbol{l}^T \boldsymbol{X} = 0.$$

We will briefly elucidate the properties of such a quadric by means of the hyperboloid of one sheet which is its non-degenerate shape. The hyperboloid of one sheet (Fig. 6) is a doubly ruled surface and has two families of generators. Each family has two main generators passing through the main vertices (the intersections of the ellipse with its major axis) of the smallest ellipse (the ellipse satisfying $x^2/a^2 + y^2/b^2 = 1$ if the hyperboloid has the standard form $x^2/a^2 + y^2/b^2 - z^2/c^2 = 1$). If all the scene points lie on a hyperboloid of one sheet then this hyperboloid is an ambiguity-critical surface [58, 59] iff

- It is rectangular (every plane perpendicular to the main generator intersects the surface in a circle) and
- The two optical centers $\boldsymbol{o}$ and $\boldsymbol{a}$ lie on generators $g1$ and $g2$ symmetric to the main generator.

Suppose now that the two motion solutions differ only by first-order perturbations $(\delta\boldsymbol{\Omega}, \delta\boldsymbol{a})$:

$$U = I + [\delta\boldsymbol{\Omega}]_\times \qquad \boldsymbol{b} = \boldsymbol{a} + \delta\boldsymbol{a}, \quad (32)$$

where $[\delta\boldsymbol{\Omega}]_\times$ is the antisymmetric matrix for the vector $\delta\boldsymbol{\Omega}$. We substitute the above terms in (30) and after cancellation of the terms second-order in $(\delta\boldsymbol{\Omega}, \delta\boldsymbol{a})$ we obtain

$$(\delta\boldsymbol{\Omega} \times \boldsymbol{X})^T (\boldsymbol{X} \times \boldsymbol{a}) + (\boldsymbol{a} \times (\boldsymbol{a} \times \delta\boldsymbol{\Omega} + \delta\boldsymbol{a}))^T \boldsymbol{X} = 0. \quad (33)$$

which has also the form $\boldsymbol{X}^T M \boldsymbol{X} + \boldsymbol{l}^T \boldsymbol{X} = 0$ where

$$M = \frac{1}{2}(\boldsymbol{a}\delta\boldsymbol{\Omega}^T + \delta\boldsymbol{\Omega}\boldsymbol{a}^T) - \boldsymbol{a}^T\delta\boldsymbol{\Omega} I \quad (34)$$

$$\boldsymbol{l} = \boldsymbol{a} \times (\boldsymbol{a} \times \delta\boldsymbol{\Omega} + \delta\boldsymbol{a}). \quad (35)$$

From the form of the M-matrix it can be easily proved that the quadric is still rectangular[b]. We will prove that the optical centers $\boldsymbol{o}$ and $\boldsymbol{a}$ lie on the main generator of the quadric.

The main generator of the quadric passes through the main vertex of the smallest intersection ellipse. The center of the quadric is given by [60]

$$\boldsymbol{c} = -\frac{1}{2} M^{-1} \boldsymbol{l}. \quad (36)$$

The main axis of the considered ellipse is parallel to the second eigenvector of the matrix M

$$\boldsymbol{u}_2 = \frac{\boldsymbol{a} \times \delta\boldsymbol{\Omega}}{\|\boldsymbol{a} \times \delta\boldsymbol{\Omega}\|}$$

[a] See [55] *"Because of his [54] particular formulation of the problem, the relationship between his results and those presented here is hard to establish."*

[b] The algebraic definition for rectangularity of a quadric is that the middle eigenvalue of M is equal to the sum of the other two.

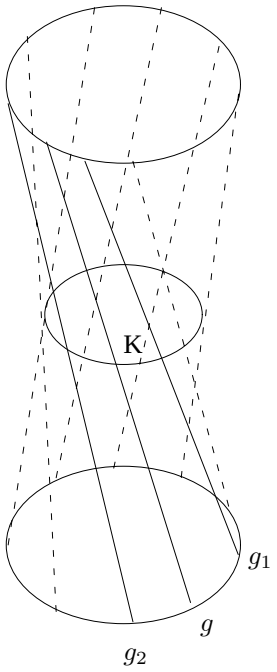


Figure 6: The hyperboloid of one sheet with three lines belonging to the same generator family. The line g through the main vertex of the smallest ellipse is the main generator of the family

corresponding to the eigenvalue $\lambda_2 = -\boldsymbol{a}^T \delta\boldsymbol{\Omega}$. The length of the semi-axis is

$$\sqrt{c/\lambda_2} \qquad \text{with} \qquad c = \frac{1}{4}\boldsymbol{l}^T M^{-1} \boldsymbol{l}.$$

Hence, the main vertices have position vectors

$$\boldsymbol{c} \pm \sqrt{\frac{c}{\lambda_2}}\boldsymbol{u}_2.$$

After a tedious calculations it can be proved that the line

$$\boldsymbol{c} \pm \sqrt{\frac{c}{\lambda_2}}\boldsymbol{u}_2 + \mu\boldsymbol{a}$$

is the main generator. On the other hand the line $\lambda\boldsymbol{a}$ joining the two optical centers is a generator since $\lambda^2 \boldsymbol{a}^T M \boldsymbol{a} + \lambda \boldsymbol{l}^T \boldsymbol{a} = 0$ for every λ. Because no two lines of a generator family are parallel to each other the line $\lambda\boldsymbol{a}$ is identical to the main generator. Hence, if the two solutions differ by a first-order perturbation the two optical centers lie on the main generator. The instability-critical quadric (33) is identical with the quadric obtained by Horn [54] who studied which scene point configurations let the epipolar constraint ${\boldsymbol{x}_1}^T(\boldsymbol{a} \times R\boldsymbol{x}_2)$ unaffected by a first-order perturbation $(\delta\boldsymbol{\Omega}, \delta\boldsymbol{a})$. We, thus, made the required link between instability- and ambiguity-critical surfaces.

The degenerate forms of a hyperboloid of one sheet are the hyperbolic paraboloid, the cone, the elliptic cylinder, and the plane pair. We will look into the case of the elliptic cylinder. We obtain a cylinder in the special case of $\boldsymbol{a} \parallel \delta\boldsymbol{\Omega}$. The generators of a cylinder are all parallel to its axis $\boldsymbol{a}$. Thus, we obtain an unstable solution if an observer is moving on a cylinder parallel to the axis and is looking the other side of the cylinder. This is a very realistic situation in case of a small field of view because a small part of a scene can be often approximated by a quadratic cylinder patch. Since the translation perturbation $\delta\boldsymbol{a}$ is perpendicular to the translation $\boldsymbol{a}$ due to $\|\boldsymbol{a}\| = 1$ we obtain a second slightly different solution including a rotation and a translation around axes $(\delta\boldsymbol{\Omega}, \delta\boldsymbol{a})$ perpendicular to each other. We gave, thus, yet another explanation for the confounding between translation and rotation.

3. Statistical Bias

The goal of estimation is to recover a set of unknown parameters with as little deviation from the ground truth as possible and in practice we need several ways to qualitatively describe the nature of the deviation that we can expect. We concentrate again on two of them, the *bias* and the *variance*. The bias is the difference between the expected value of the estimate and the ground truth and the variance is the expected value of the squared difference between the estimate and the ground truth. It seems obvious that in most cases reducing the bias will reduce the variance and the opposite but this is increasingly incorrect as our estimators become better and more complicated so we move away from straightforward intuition. It is not uncommon to have unbiased solutions with high variance, or biased solution to outperform unbiased ones in terms of variance. We could restrict ourselves to linear estimators with Gaussian distribution and preserve the monotone relation between bias and variance but this could be hardly useful for structure estimation using discrete matches where even the so called *linear solutions* include nonlinear steps.

T he bias itself can be a very useful property and worth striving to include among the advantages of an algorithm. It is not without caveats though. Assume that we want to recover the size of a square in three dimensional space. We use an algorithm to obtain an unbiased estimate of the length of the sides of the square

$$E\{x^*\} = x$$

where x^* is the estimate and x is the actual length of the side of the square. But using the same algorithm we cannot get an unbiased estimate of the area x^2 by simply squaring the result because

$$E\{x^{2^*}\} = x^2 + \sigma^2 \neq x^2$$

where σ is the standard deviation of x. In other words the bias is a very fragile property and it might not be always easy to obtain estimators whose bias is exactly zero although in most cases simple corrections are sufficient.

The estimation of the rotation component of a general motion poses similar problems. If one designs an unbiased estimator of the Euler angles of the motion then the same estimator cannot be simply translated to one that computes the Rodriguess parameters. Even the translation component poses difficulties depending on whether is reported as a focus of expansion or a unit vector.

Then one cannot prove that a certain estimator is unbiased without reference to the exact representation of the required result because if a certain representation is unbiased we cannot infer that an equivalent one is also unbiased. On the other hand the bias that appears when we simply change representation is very often too small to be observable by plotting or visualizing the data. If the bias is due to deficiencies in the formulation of the estimator then in several cases it is very noticeable by simple inspection of the data, which is what happened to the structure from motion problem [8, 9].

Several of the oldest techniques to recover the motion parameters from point matches in two frames involve the minimization of the sum of the squared residuals of the epipolar constraint

$$\boldsymbol{p}_{2_i}{}^T E \boldsymbol{p}_{1i} = 0$$

where $\boldsymbol{p}_{1_i}$ and $\boldsymbol{p}_{2_i}$ are the images of the i^th point in the first and second image, the matrix E is

$$E = \begin{pmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{pmatrix} R$$

with $\boldsymbol{T} = (T_x, T_y, T_z)^T$. The minimization of

$$\sum_i \left(\boldsymbol{p}_{2_i}{}^T E \boldsymbol{p}_{1i}\right)^2$$

is fairly easy and there are several solutions in the literature [6, 61] but it is heavily biased in the presence of noise. To see intuitively why we can rewrite the epipolar constraint as

$$\boldsymbol{p}_{2_i}{}^T E \boldsymbol{p}_{1i} = \boldsymbol{p}_{2_i}{}^T (\boldsymbol{T} \times (R\boldsymbol{p}_{1i}))$$

It is easy to see that the cross product introduces the sine of the angle between vectors $\boldsymbol{T}$ and $R\boldsymbol{p}_{1i}$ as a factor. This will make the minimization to prefer vectors $\boldsymbol{T}$ that are closer to the center of gravity of the bundle of points $R\boldsymbol{p}_{1i}$ so that the sine and hence the residual of the epipolar is smaller [9].

The magnitude of the bias and the suspicious presence of the sine in the form of the epipolar suggest that there is some oversimplification in the way we handled this minimization. We can use Maximum Likelihood method from statistics and solve the problem as follows [9]. Assume that every world point $\boldsymbol{P}_{1i}$ is equal to $\boldsymbol{P}_{1i} = Z_i\boldsymbol{p}_{1i}$, where $\boldsymbol{p}_{1i}$ is the image point vector and Z_i is the unknown scalar depth (which is equal to the z component in planar images). If we move the point $\boldsymbol{P}_{1i}$ by a rotation R and a translation $\boldsymbol{T}$ we get

$$\boldsymbol{P}_{2i} = Z_i R\boldsymbol{p}_{1i} + \boldsymbol{T}$$

If we project this vector on the second camera we get

$$\boldsymbol{p}_{2i}'' = \frac{\boldsymbol{P}_{2i}}{\hat{\boldsymbol{z}}^T\boldsymbol{P}_{2i}}$$

which in the presence of noise does not coincide with the projection of the same point $\boldsymbol{p}_{2i}$ on the second camera. The noise is the distance between these two vectors $\boldsymbol{n}_i = \boldsymbol{p}_{2i}'' - \boldsymbol{p}_{2i}$. A simple application of the the maximum likelihood method can show that we need to minimize the weighted least squares of the noise for all points

$$\sum_i {\boldsymbol{n}_i}^T {\Sigma_i}^+ \boldsymbol{n}_i \rightarrow min.$$

where ${\Sigma_i}^+$ is the pseudoinverse of a covariance matrix of suitable form (e.g. for planar images only the top left two by two submatrix is important). We can eliminate Z_i which is one unknown per pixel by setting it equal to

$$Z_i = \frac{(\boldsymbol{p}_{2i} \times \boldsymbol{T})^T \Sigma_i (\boldsymbol{T} \times R\boldsymbol{p}_{1i})}{(\boldsymbol{T} \times R\boldsymbol{p}_{1i})^T \Sigma_i (R\boldsymbol{p}_{1i} \times \boldsymbol{p}_{2i})}.$$

After the elimination the expression we have to minimize is

$$\sum_i {\boldsymbol{n}_i}^T {\Sigma_i}^+ \boldsymbol{n}_i = \sum_i \frac{({\boldsymbol{p}_{2i}}^T E\boldsymbol{p}_{1i})^2}{(E\boldsymbol{p}_{1i})^T \Sigma_i E\boldsymbol{p}_{1i}}.$$

which is a much more difficult function to minimize than the simple sum of squared epipolars. One word of caution here. The expression in the denominator is not a weight that increases the importance of the points close to the focus of expansion (for these points the expression is close to zero). It contains the unknowns so it cannot be factored out the same way as weights and has to remain part of the minimization. And of course the minimization procedure is more complicated as a result.

As can be seen from the experimental data in [9] the error distribution does not look biased. The question remains though. Is it biased? Just because it does not look biased it does not mean it is not. And in fact it is not unbiased. The failure of such a bias to manifest itself in the experimental data indicates that this is small and certainly it is smaller than the one we get from minimizing the squared sum of the residuals of the epipolar.

4. Using Lines and Points

Straight line matching does not suffer from the aperture problem, where the one of the two components of the flow field is impossible to estimate if one looks through a sufficiently small aperture. But the aperture problem does not appear everywhere in the image. There are feature points in the image that do not suffer from the aperture problem (e.g. corners and local maxima). If one wants to make use of these feature points then information from the line matching and the point matching should be used at the same time. Such an algorithm needs of course 3 frames, otherwise the lines do not offer any real constraint.

The idea of using 3 frames for points without lines has some other advantages of its own. First fewer points are needed to obtain a finite number of solutions (four instead of five, but for linearized solution we

need nine instead of eight). This is because there is the extra constraint that the structure obtained from the first and second frame is the same as the one obtained from the first and third. Second the ambiguous configurations are fewer. Configurations that are ambiguous in the first pair of frames are either unambiguous in the second pair or the spurious solution is different [23].

Assume that a pont P in 3-D is projected to points $\boldsymbol{p}_1$, $\boldsymbol{p}_2$ and $\boldsymbol{p}_3$ on the three frames. The motion between frame 1 and frame 2 is R_1 and $\boldsymbol{T}_1$ for rotation and translation respectively. The motion between frame 1 and 3 is similarly R_2 and $\boldsymbol{T}_2$. The image points are related by

$$Z_2\boldsymbol{p}_2 = Z_1R_1\boldsymbol{p}_1 + \boldsymbol{T}_1$$
$$Z_3\boldsymbol{p}_3 = Z_1R_2\boldsymbol{p}_1 + \boldsymbol{T}_2$$

where Z_i is the unknown depth of the points in the corresponding coordinate system of the frame. We eliminate Z_2 and Z_3 by taking the cross product of the two equations by $\boldsymbol{p}_2$ and $\boldsymbol{p}_3$ respectively.

$$0 = Z_1\boldsymbol{p}_2 \times (R_1\boldsymbol{p}_1) + \boldsymbol{p}_2 \times \boldsymbol{T}_1$$
$$0 = Z_1\boldsymbol{p}_3 \times (R_2\boldsymbol{p}_1) + \boldsymbol{p}_3 \times \boldsymbol{T}_2$$

We can eliminate Z_1 by rearranging the terms so that Z_1 appears in the left hand side for the one equation and the right hand side of the other equation

$$Z_1[\boldsymbol{p}_2]_\times R_1\boldsymbol{p}_1 = -[\boldsymbol{p}_2]_\times \boldsymbol{T}_1$$
$$[\boldsymbol{p}_3]_\times \boldsymbol{T}_2 = -Z_1[\boldsymbol{p}_3]_\times R_2\boldsymbol{p}_1$$

and then taking the outer product we get the following constraint

$$[\boldsymbol{p}_2]_\times (R_1\boldsymbol{p}_1\boldsymbol{T}_2^T)[\boldsymbol{p}_3]_\times = [\boldsymbol{p}_2]_\times \boldsymbol{T}_1(R_2\boldsymbol{p}_1)^T\boldsymbol{p}_3$$

where

$$[\boldsymbol{p}_2]_\times = \begin{pmatrix} 0 & -z_2 & y_2 \\ z_2 & 0 & -x_2 \\ -y_2 & x_2 & 0 \end{pmatrix}$$

with $\boldsymbol{p}_2 = (x_2, y_2, z_2)^T$. If we define

$$K = \boldsymbol{T}_1(R_2\hat{\boldsymbol{x}})^T - (R_1\hat{\boldsymbol{x}})\boldsymbol{T}_2{}^T$$
$$L = \boldsymbol{T}_1(R_2\hat{\boldsymbol{y}})^T - (R_1\hat{\boldsymbol{y}})\boldsymbol{T}_2{}^T$$
$$M = \boldsymbol{T}_1(R_2\hat{\boldsymbol{z}})^T - (R_1\hat{\boldsymbol{z}})\boldsymbol{T}_2{}^T$$

where $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{z}}$ are the unit vectors along the corresponding axes, then the constraint becomes

$$[\boldsymbol{p}_2]_\times (x_1K + y_1L + z_1M)[\boldsymbol{p}_3]_\times = [0]$$

where (x_1, y_1, z_1) are the components of $\boldsymbol{p}_1$. This is a matrix equation that is equivalent to 9 scalar equations but only three of them are independent.

The above equation provides a set of constraints for points. The known data are the points $\boldsymbol{p}_{1_i}$ and the unknowns are the rotation and translation parameters which are hidden in the three matrices K, L and M. The constraints for the lines is the vector equation

$$\begin{pmatrix} \boldsymbol{\epsilon}_2{}^T K\boldsymbol{\epsilon}_3 \\ \boldsymbol{\epsilon}_2{}^T L\boldsymbol{\epsilon}_3 \\ \boldsymbol{\epsilon}_2{}^T M\boldsymbol{\epsilon}_3 \end{pmatrix} \times \boldsymbol{\epsilon}_1 = \vec{0}$$

where $\boldsymbol{\epsilon}_i$ is the normal vector line representation in frame i. Such a vector passes through the origin and is normal to the image line. Known are the normal line vectors and the unknowns are again the matrices K, L and M that contain the motion parameters. The constraints from both points and lines can be included in the same least squares estimation to take advantage of all the information.

But this least squares problem corresponds to the direct least squares minimization for the epipolar residual that was shown above to be the reason for the high bias. The question arises if we can follow the same path and find the maximum likelihood estimator for the points and lines as well. It turns out that we cannot do exactly the same because we need to solve a fourth degree polynomial equation that although it has a closed form solution, it is no more useful than a numerical one due to its complexity.

5. Statistically Correlated Noise

In the previous paragraphs we dealt mainly with data that were statistically independent. This is a convenient working assumption that is fairly accurate when we deal with point correspondences that are collected either by hand or sparsely enough to make them independent. But a computer vision system should not depend on data collected by hand and sparse data means that we leave large amount of information unused.

Furthermore, it is possible to extract more information from data corrupted by correlated noise if we know the correlation rather than assuming incorrectly that they are uncorrelated. One example is the motion of a textured patch. While any measurement of the motion of such a patch may be corrupted due to aliasing, the second order components of the motion (dilation, shear, rotation) may be still recoverable, if we can track the Fourier components in the frequency domain.

There are two distinct problems regarding correlated noise. One is the representation of the variance-covariance matrix. Such a matrix should contain the covariance for every pair of points in the image, which is very large for dense disparity data.

The other problem is that the equations cannot be solved or simplified analytically, because all the image is involved in one set of equations. The equations have to be solved numerically.

The approach to these problems is to solve for the correspondence (or flow) and for the motion in one step, by minimizing the match over a region of the intensity subject to the constraint that the correspondences satisfy the epipolar constraint. The Lagrangian equation is

$$\begin{aligned}\mathcal{L} &= \sum_{i_0,j_0} \sum_{\substack{i_0-w\le i\le i_0+w \\ j_0-w\le j\le j_0+w}} \left(\nabla I(i,j)^T \boldsymbol{u}(i_0,j_0) + I_t(i,j)\right)^2 + \\ &\quad \sum_{i_0,j_0} \lambda(i_0,j_0)\left(\boldsymbol{p}_2(i_0,j_0)^T E \boldsymbol{p}_1(i_0,j_0)\right)\end{aligned}$$

where the summation for i, j is over the whole image, i_0 and j_0 is over a small window typically 5 by 5 or 7 by 7,

$$\boldsymbol{p}_2(i_0,j_0) = \boldsymbol{p}_1(i_0,j_0) + \boldsymbol{u}(i_0,j_0)$$

$\boldsymbol{u}$ is the disparity with zero as the third element to match the size of $\boldsymbol{p}_2$, I is the intensity and I_t is the time derivative of the intensity. The first component of $\mathcal{L}$ is the metric used by Lucas and Kanade [27] and seems to produce reasonable accurate results even by itself [28]. Solving this Lagrangian equation for $\boldsymbol{u}$ and E will give a disparity field that is constrained to be one of rigid motion. But the linear equations that come out are hard to solve because the linearized system is non positive definite. Although iterative methods exist for this type of equations they are slow and unstable. In general it is better to avoid such systems. And although we are not interested in λ, we have to compute it in the process. Adding the constraint that matrix E is decomposable to rotation and translation provides an extra difficulty.

Another formulation of the problem to use a trade off parameter λ which yields a positive definite system

$$\begin{aligned}\mathcal{L} &= \sum (\nabla I(i-i_0, j-j_0)^T \boldsymbol{u}(i_0,j_0) + I_t(i-i_0, j-j_0))^2 + \\ &\quad \lambda \sum \frac{(\boldsymbol{p}_2(i_0,j_0)^T E \boldsymbol{p}_1(i_0,j_0))^2}{(E\boldsymbol{p}_1(i_0,j_0))^T \Sigma_i E \boldsymbol{p}_1(i_0,j_0)}\end{aligned}$$

which is a form of regularization but instead of trading off the problem constraints against a smoothness term, we do so against a rigidity term which in many cases offers more than stabilization: It can link the disparity estimation to the motion estimation. Unfortunately if the disparity estimation is severely affected by the aperture problem the whole process is rendered ineffective. Such methods work quite well [62] but the problem of choosing the appropriate λ and Σ_i remains. In general the value of λ has to be as high as possible since it will not have bad side effects like oversmoothing the solution, but too high can be detrimental to the round off error.

A third approach involves unconstrained minimization by changing unknowns. Instead of having the components of the flow vector field and the motion parameters as unknowns we can use the depth and the motion parameters

$$\mathcal{L} = \sum (\nabla I(i - i_0, j - j_0) \cdot \boldsymbol{u}(i_0, j_0) + I_t(i - i_0, j - j_0))^2$$

where

$$\boldsymbol{u} = \boldsymbol{p}_1 - \frac{ZR\boldsymbol{p}_1 + \boldsymbol{T}}{\hat{\boldsymbol{z}}^T(Z\boldsymbol{p}_1 + \boldsymbol{T})}$$

where Z is the z component of the point before the motion. The flow vectors are simply intermediate variables. A disadvantage of this approach is that it has more difficult nonlinearities to handle but this can be weighted against the advantage of fewer unknowns and the ability to incorporate information about the depth that comes from another source (previous frames, sonar, known models etc). The linearized equations that have to be solved are also positive definite.

All these approaches share something with the ideas behind direct methods [17] and correspondenceless methods [18, 63] for motion recovery that avoids the intermediate step of establishing correspondence or any other kind of disparity field. What they share is the one step estimation of the motion. But they differ in the kind of input they expect. Direct methods attempt to recover the motion using less information, either constraints that only give the normal component of the motion or the distribution of points without one to one correspondence established between them. Methods presented above assume that there is sufficient information in the image to recover deformation. And while it is true that the constraints on the disparity are weak in large areas on the image, using some of the best performing algorithms [28] both components of flow can be recovered in significant portions of the image and these are enough to recover the motion. The areas that simply do not have enough constraints for both components of flow, do not contribute to the recovery of motion or even to the recovery of depth if no component of flow can be computed reliably.

6. The future

The application that has driven the research on structure from motion is the autonomous navigation where scene reconstruction would make the problem simpler, although it has been argued that autonomous navigation could be successful without it. Nevertheless an efficient structure from motion module would be useful for the navigation itself or the other tasks of the autonomous robots. A parallel can be drawn from marine navigation where an accurate map and the exact location of the vessel is not needed in safe waters near the coast. But for offshore navigation or in unsafe waters not only the exact location on the map is required but a radar as well.

But the recent demand for multimedia, high quality graphics and virtual reality, has opened new opportunities for applications of structure from motion. One problem that structure from motion can help is viewpoint interpolation where given the image of a scene from a set of different viewpoints one wants the image of the scene from a novel viewpoint. Although with today's technology depth cannot be recovered accurately, it is at least consistent with the gray scale (or color) structure of the scene, so if the novel viewpoint is not very far from the already existing ones one can compute the image from that viewpoint. Another application is interactive video editing where two or three dimensional tracking can help map the editing of one frame to the next. There are numerous other applications for scene reconstruction techniques that will emerge sooner or later. The point is that general purpose scene reconstruction has much broader importance outside what has become tradition in vision research. The presented chapter not only provides us with the reasons of sensitivity but gives us also a unifying way of its treatment by applying solid statistical techniques.

On the other hand, if we follow the paradigm of active and / or purposive vision the noise sensitivity treatment provides us with guidelines on how to design a purposive algorithm. Given a specific task like the estimation of object translation and the pursuing of the object we may find out that an object centered representation eliminates the translation-rotation confounding – in case of a small field of view of object tracking – because the direction of least uncertainty in the space of unknowns is now the lateral translation of the object. In the task of ego-translation estimation using fixation we can exploit the results on the important line in the space of translation directions. In particular, it was proved that the use of the polar

transformation [64] decouples the azimuth from the polar angle of the translation direction, the former of which shown already to be insensitive to noise. We hope that this chapter will give further insight towards designing not only computationally inexpensive but also robust general as well as special purpose algorithms.

Acknowledgements

We sincerely thank our advisors Professors Hans-Hellmut Nagel and Yannis Aloimonos who introduced us to the subject and motivated us through the entire course of this research. Their contribution through discussions, comments, and steady encouragement is invaluable.

References

[1] H.C. Longuet-Higgins. A computer program for reconstructing a scene from two projections. *Nature*, 293(11):133–135, 1981.

[2] R. Y. Tsai and T. S. Huang. Uniqueness and estimation of 3-d motion parameters of rigid bodies with curved surfaces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6:13–27, 1984.

[3] H.C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. *Proc. Royal Society of London*, B208:385–397, 1980.

[4] A.M. Waxman and K. Wohn. Contour evolution, neighborhood deformation, and global image flow: planar surfaces in motion. *Intern. Journal of Robotics Research*, 4(3):95–108, 1985.

[5] C. C. Slama, C. Theurer, , and S. W. Henriksen. *Manual of Photogrammetry*. American Society of Photogrammetry, 1980.

[6] B.K.P. Horn. Relative orientation revisited. *Journal Opt. Soc. Am.*, A8:1630–1638, 1991.

[7] C. Jerian and R. Jain. Polynomial methods for structure from motion. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12:1150–1165, 1990.

[8] M.E. Spetsakis and Y. Aloimonos. Optimal visual motion estimation: A note. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14:959–964, 1992.

[9] M. Spetsakis. Models of statistical visual motion estimation. *Computer Vision, Graphics, and Image Processing*, 60:300–312, 1994.

[10] J. Weng, N. Ahuja, and T.S. Huang. Optimal motion and structure estimation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15:864–884, 1993.

[11] K. Daniilidis and H.-H. Nagel. Analytical results on error sensitivity of motion estimation from two views. *Image and Vision Computing*, 8:297–303, 1990.

[12] K. Daniilidis and H.-H. Nagel. The coupling of rotation and translation in motion estimation of planar surfaces. In *IEEE Conf. on Computer Vision and Pattern Recognition 1993*, pages 188–193, New York, NY, June 15-17, 1993.

[13] G. Adiv. Inherent ambiguities in recovering 3-d motion and structure from a noisy flow field. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11:477–489, 1989.

[14] G.-S.J. Young and R. Chellappa. Statistical analysis of inherent ambiguities in recovering 3-d motion from a noisy flow field. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14:995–1013, 1992.

[15] K. Kanatani. Unbiased estimation and statistical analysis of 3-d rigid motion from two views. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15:37–50, 1993.

[16] K. Kanatani. *Geometric Computation for Machine Vision*. Oxford University Press, Oxford, UK, 1993.

[17] S. Negahdaripour and B.K.P. Horn. Direct passive navigation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9:168–176, 1987.

[18] A. Basu and J. Aloimonos. A robust, correspondenceless translation-determining algorithm. *Intern. Journal of Robotics Research*, 9(5):35–59, 1990.

[19] D.J. Heeger and A.D. Jepson. Subspace methods for recovering rigid motion i: Algorithm and implementation. *International Journal of Computer Vision*, 7:95–117, 1992.

[20] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 8:137–154, 1992.

[21] Y. Liu and T.S. Huang. Estimation of rigid body motion using straight line correspondences. *Computer Vision, Graphics, and Image Processing*, 43:37–52, 1988.

[22] M. Spetsakis and J. Aloimonos. Structure from motion using line correspondences. *International Journal of Computer Vision*, 4:171–183, 1990.

[23] M.E. Spetsakis. A linear algorithm for point and line- based structure from motion. *CVGIP: Image Understanding*, 56:230–241, 1992.

[24] J.(Y.) Aloimonos, I. Weiss, and A. Bandyopadhyay. Active vision. *International Journal of Computer Vision*, 1:333–356, 1988.

[25] J. Aloimonos. Purposive and qualitative active vision. In *Proc. Int. Conf. on Pattern Recognition*, pages 346–360, Atlantic City, NJ, June 17-21, 1990.

[26] K.N. Kutulakos and C.R. Dyer. Recovering shape by purposive viewpoint adjustment. *International Journal of Computer Vision*, 12:113–136, 1994.

[27] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *DARPA Image Understandig Workshop*, pages 121–130, 1981.

[28] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12:43–78, 1994.

[29] B.K.P. Horn and B.G. Schunk. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

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

[31] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2:283–310, 1989.

[32] A. Singh. *Optic Flow Computation*. IEEE Computer Society Press, Los Alamitos, CA, 1991.

[33] P. Werkhoven and J.J. Koenderink. Extraction of motion parallax structure. *Biological Cybernetics*, 63:185–191, 1990.

[34] M. Campani and A. Verri. Optical flow estimation using discontinuity conforming filters. In *Proc. Int. Conf. on Computer Vision*, pages 22–26, Osaka, Japan, Dec. 4-7, 1990.

[35] D. J. Fleet, A. D. Jepson, and M. R. M. Jenkin. Phase-based disparity measurement. *CVGIP: Image Understanding*, 53(2), 3 1991.

[36] M. E. Spetsakis. Optical flow estimation using discontinuity conforming filters. In *British Machine Vision Conference, York, UK*, 1994.

[37] A. Jepson and D.J. Heeger. A fast subspace algorithm for recovering rigid motion. In *Proc. IEEE Workshop on Visual Motion*, pages 124–131, Princeton, NJ, Oct. 7-9, 1991.

[38] A.D. Jepson and D.J. Heeger. Subspace methods for recovering rigid motion ii: Theory. Technical Report RBCV-TR-90-36, University of Toronto, 1990.

[39] K. Kanatani. 3-d interpretation of optical flow by renormalization. *International Journal of Computer Vision*, 11(267-282), 1993.

[40] J.L. Barron, A.D. Jepson, and J.K. Tsotsos. The feasibility of motion and structure from noisy time-varying image velocity information. *International Journal of Computer Vision*, 5:239–269, 1990.

[41] C.-H. Lee. Time-varying images: The effect of finite resolution on uniqueness. *CVGIP: Image Understanding*, 54:325–332, 1991.

[42] S.J. Maybank. *A theoretical study of optical flow*. PhD thesis, University of London, November 1987.

[43] H.W. Sorenson. *Parameter Estimation, Principles and Problems*. Marcel Dekker, New York and Basel, 1980.

[44] H.C. Longuet-Higgins. The visual ambiguity of a moving plane. *Proc. Royal Society of London*, B223:165–175, 1984.

[45] B. Friedland. *Control System Design*. McGraw-Hill, New York, 1986.

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

[47] M.E. Spetsakis and J. Aloimonos. Optimal motion estimation. In *Proc. IEEE Workshop on Visual Motion*, pages 229–237, Irvine, CA, March 20-22, 1989.

[48] J. Weng, T.S. Huang, and N. Ahuja. Optimal motion and structure estimation. In *IEEE Conf. Computer Vision and Pattern Recognition*, pages 144–152, San Diego, CA, June 4-8, 1989.

[49] J. Aisbett. An iterated estimation of the motion parameters of a rigid body from noisy displacement vectors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12:1092–1098, 1990.

[50] A. Bruss and B.K.P. Horn. Passive navigation. *Computer Vision, Graphics, and Image Processing*, 21:3–20, 1983.

[51] T. Lawton. Processing translational motion sequences. *Computer Vision, Graphics, and Image Processing*, 22:116–144, 1983.

[52] G. Adiv. Determining 3-d motion and structure from optical flow generated by several moving objects. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 7:384–401, 1985.

[53] S. Maybank. *Theory of Reconstruction from Image Motion*. Springer-Verlag, Berlin et al., 1993.

[54] B.K.P. Horn. Relative orientation. *International Journal of Computer Vision*, 4:59–78, 1990.

[55] S. Negahdaripour. Multiple interpretations of the shape and motion of objects from two perspective images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12:1025–1039, 1990.

[56] W. Hofmann. *Das Problem der "gefährlichen Flächen" in Theorie und Praxis*. Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften, Reihe C, Heft 3, München, 1953.

[57] K. Daniilidis. On the error sensitivity in the recovery of object descriptions and relative motions from image sequences. Doctoral dissertation, Department of Informatics, University of Karlsruhe, Germany, July 1992. in German.

[58] J. Krames. Zur Ermittlung eines Objektes aus zwei perspektiven - Ein Beitrag zur Theorie der "gefährlichen örter". *Monatshefte für Mathematik und Physik*, 49:327–354, 1940.

[59] W. Wunderlich. Zur Eindeutigkeitsfrage der Hauptaufgabe der Photogrammetrie. *Monatshefte für Mathematik und Physik*, 50:151–164, 1941.

[60] G.A. Korn and T.M. Korn. *Mathematical Handbook for Scientists and Engineers*. McGraw-Hill, New York, 1968.

[61] M.E. Spetsakis and J. Aloimonos. Optimal computing of structure from motion using point correspondences. In *Proc. Int. Conf. on Computer Vision*, pages 449–453, Tampa, FL, Dec. 5-8, 1988.

[62] H. Ando. Dynamic reconstruction of 3d structure and 3d motion. In *Proc. IEEE Workshop on Visual Motion*, pages 101–110, Princeton, NJ, Oct. 7-9, 1991.

[63] J. Aloimonos and J.-Y. Herve. Correspondenceless stereo and motion: planar surfaces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12:504–510, 1990.

[64] K. Daniilidis. Attentive visual motion processing: computations in the log-polar plane. *Computing, Archives in Informatics and Numerical Mathematics*, 1995. To appear in the special issue on Theoretical Foundations of Computer Vision.