The Coupling of Rotation and Translation in Motion Estimation of Planar Surfaces

Konstantinos Daniilidis[†*]
[†]Institut für Algorithmen und Kognitive Systeme
Fakultät für Informatik, Universität Karlsruhe (TH)

Hans-Hellmut Nagel[†‡]
[‡]Fraunhofer-Institut für Informations- und Datenverarbeitung (IITB), Karlsruhe

Abstract

This paper studies the error sensitivity in the estimation of the 3D-motion and the normal of a planar surface from an instantaneous motion field. We use the statistical theory of the Cramer-Rao lower bound for the error covariance in the estimated motion and structure parameters which enables the derivation of results valid for any unbiased estimator under the assumption of Gaussian noise in the motion field. The obtained lower-bound-matrix is studied analytically with respect to the measurement noise, size of the field of view and the motion-geometry configuration. The main result of this analysis is the coupling between translation and rotation which is exacerbated if the field of view and the slant of the plane become smaller and the deviation of the translation from the viewing direction becomes larger. By-products of this study are the relationships of the uncertainty bounds for every unknown motion parameter to the angle between translation and the plane-normal, the size of the field of view, the distance from the perceived plane and the translation magnitude.

1 Introduction

Three-dimensional motion and structure estimation from monocular image sequences has been studied extensively in the fields of computer vision, perceptual psychology and neurobiology. Many computational theories have been developed and many algorithms have been proposed in order to enrich mobile robots with the ability to interact in a changing environment. It turned out that this general problem formulation suffers from the existence of more than one solution – the ambiguity problem – and the high sensitivity to measurement noise. Recently new problem formulations have been stated that follow the latest paradigms of qualitative, purposive and/or active vision in order to overcome the ambiguity and the sensitivity problem. Nevertheless, questions with regard to these problems remain still open and our effort is to find analytical answers in order to guarantee when a proposed technique will exhibit a stable behavior and when not.

*Correspondence to K. Daniilidis, Computer Science Institute, Kiel University, Preusserstr. 1-9, D-2300 Kiel, Germany, email: kd@informatik.uni-kiel.dbp.de

We will follow the classical two-steps way in 3D-motion estimation from a monocular image sequence. The first consists of the computation of feature correspondences (discrete case) or optical flow vectors (continuous case) induced by the relative motion between the camera and the environment. What we can measure in the image are apparent shifts or velocities of gray-value structures which are approximations to the geometrically defined displacement or velocities of the projections of three-dimensional features. We call the latter motion field in contrast to the former which we call optical flow field [33]. Our analysis refers to the continuous case and the existence of a dense motion field. The second step – as of now we will consider only the continuous case – consists in the estimation of translational and angular velocities as well as of the distances to the points in the scene. The error sensitivity of the second step to the error in the measurements of the first step is the focus of our study. For a survey of algorithms for 3D-motion and structure computation the reader is referred to [14] for the discrete case and [2] for the discrete as well as the continuous case.

For every point in the image associated with a motion field vector we must, in general, introduce its distance to the corresponding 3D-point as an unknown. The study of the sensitivity in the structure estimates and of the dependence of the estimation error on the structure thereby becomes analytically intractable. In order to reduce the number of unknowns we restrict our analysis to the case of a planar surface. We, thus, can describe the structure by only two unknowns which give the direction of the plane's normal. The distance of the plane from the camera is coupled with the translation magnitude due to the well known scale-ambiguity. Furthermore, the motion field obtains a very special form: It is quadratic with respect to the image point positions and can be fully described by eight parameters. The functional dependence of the motion field measurements to these parameters is linear whereas these parameters are bilinear in the translational velocity and the normal and linear in the angular velocity.

Solutions for the 3D-motion and the normal of a planar surface have been proposed by [18, 5, 34, 28, 25, 15] for the continuous case and by [32, 7, 35] for the discrete case. All approaches are based on the solution of a cubic equation derived either directly from the motion field equations or as the characteristic equation of a 3×3 symmetric matrix. Two solutions for the motion parameters and the normal exist if the translation is not perpendicular to the plane. This twofold ambiguity has been proved repeatedly by [16, 10, 18, 20, 19, 24].

The case of planar surfaces in motion estimation is of special interest regarding applications. Navigational tasks like autonomous vehicle driving – both outdoor and indoor – and aircraft landing include the interpretation of a motion field induced by the motion of the camera relative to a planar ground. The special quadratic form of this field allows the detection of obstacles as well as of other moving scene components. In assembly operations the case of polyhedral objects is very common. Robot manipulators should be enabled to trace a trajectory towards the planar face of an object using the motion field recorded by a camera on the gripper.

The sensitivity in motion estimation has been an object of experimental as well as of analytical investigations. [21] and later [11] proved that the minima of the error surface lie in the neighborhood of a particular line on the unit sphere of translation directions if the surface is sufficiently nonplanar. This line connects points in the unit sphere corresponding to the translation and the viewing direction. [4] and [36, 35] show how the output-error amplification and the error variance, respectively, can be computed as a function of the error in the input data, however, this dependency is not given in closed-form but as a procedure. By means of synthetic data they show that a large field of view, a large ratio of translation magnitude to distance from the moving object and a translation in the optical axis direction contribute to robustness in the motion estimates. The role of the translation direction and the geometric meaning of the error metric in use has been pointed out by [27, 30, 3, 13] and explicitly proved by [6]. [17] shows by means of concrete numerical examples that measurement noise induced only by the finite image resolution can cause a relative motion error up to 10%. Linear algorithms for the discrete case are extremely sensitive to noise as reported by [31, 36]. [12] studied the surface and motion configurations that cause a sensitivity represented by a quadratic ascent in the error function after a linear perturbation in the unknown motion parameters.

Planar surfaces in motion and the associated sensitivity in estimation have been studied by [1, 37]. [1] proved that motion fields induced by different translations and normals of the planar surface deviate from each other only in the quadratic terms. This deviation is negligible if the field of view and the plane's slant are small as well as if the translation magnitude is small compared to the distance from the object. [1] as well as [5, 22] pointed out that translation can be distinguished from rotation only be means of the slant components appearing in the first and second order terms of the motion field. We are going to show the same fact by analytically studying the lower bound of the error covariance for the velocities as well as the normal of the plane. The same technique is used by [37] who derive many results common with ours. Our contribution consists in the investigation of the interaction between translation and rotation through the analysis of the uncertainty directions – principal axes of the error ellipsoids – and departs from the analysis of [37] who inspect only the diagonal elements – i.e. the variances – of the motion parameters. Due to the complexity of the expressions derived with help of the MAPLE symbolic package we illustrate the uncertainty magnitudes and directions as function plots of the translation and normal direction. Furthermore, we give the explicit dependence of the covariance matrix on the angle between translation and the normal and we analyze the sensitivity of the normal. [37] consider only the case of the focus of expansion lying in the area of the projected moving object. Thus, they exclude the case of a relative translation to the environment parallel to the image plane and disable the analysis of the transition from a translation parallel to the viewing direction to a translation perpendicular to the viewing direction. On the other side we exclude the case of a planar surface parallel to the optical axis. We can model this case only as a limit.

In the next two sections we will outline the problem formulation and the theory of the Cramer-Rao lower bounds. Then we will compute the lower bounds and study the directions of uncertainty in dimension-reduced parameter spaces. In the last section we analyze the sensitivity in the estimation of the normal.

2 Motion field of a planar surface

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. The velocity $\dot{\boldsymbol{X}}$ of this point is given by

$$\dot{\boldsymbol{X}} = \boldsymbol{v} + \boldsymbol{\omega} \times \boldsymbol{X}. \tag{1}$$

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 $\boldsymbol{\omega}$.

We choose the origin as the center of projection and the z-axis – with $\hat{\boldsymbol{z}}$ representing its direction – as the optical axis. We assume that the focal length is unity, hence the perspective projection equation reads $\boldsymbol{x} = \boldsymbol{X}/\hat{\boldsymbol{z}}^T\boldsymbol{X}$ where $\boldsymbol{x} = (x, y, 1)^T$ is the projection on the image plane of the point $\boldsymbol{X}$. After differentiating the projection equation with respect to time, we obtain the motion field vector

$$\dot{\boldsymbol{x}} = \frac{1}{\hat{\boldsymbol{z}}^T\boldsymbol{X}}\hat{\boldsymbol{z}} \times (\boldsymbol{v} \times \boldsymbol{x}) \; + \; \hat{\boldsymbol{z}} \times (\boldsymbol{x} \times (\boldsymbol{x} \times \boldsymbol{\omega})). \tag{2}$$

In order to reduce the number of the depth unknowns we sacrifice generality and assume that the perceived surface in motion is planar. As already said in the introduction, piecewise planar environments are very common in applications. 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. (2):

$$\dot{\boldsymbol{x}} = (\boldsymbol{N}^T\boldsymbol{x})(\hat{\boldsymbol{z}} \times (\boldsymbol{v} \times \boldsymbol{x})) + \hat{\boldsymbol{z}} \times (\boldsymbol{x} \times (\boldsymbol{x} \times \boldsymbol{\omega})). \tag{3}$$

The scale ambiguity becomes evident since we observe that the one-parametric family $(\boldsymbol{v}/s, s\boldsymbol{N})$ of translation-normal pairs creates the same motion field. Thus the actual number of unknowns is eight: Three for $\boldsymbol{\omega}$, four for the directions of $\boldsymbol{v}$ and $\boldsymbol{N}$ and one for the ratio of the translation magnitude to the distance from the origin to the plane, written as $\|\boldsymbol{v}\|\,\|\boldsymbol{N}\|$.

After rearranging terms we obtain

$$\dot{\boldsymbol{x}} = \hat{\boldsymbol{z}} \times (P\boldsymbol{x} \times \boldsymbol{x}) \tag{4}$$

with

$$P = \boldsymbol{v}\boldsymbol{N}^T + [\boldsymbol{\omega}]_\times, \tag{5}$$

where $[\boldsymbol{\omega}]_\times$ is the antisymmetric matrix with the property $[\boldsymbol{\omega}]_\times\boldsymbol{x} = \boldsymbol{\omega} \times \boldsymbol{x}$. The elements of the matrix P are not independent of each other: it is easy to show that $\det(P + P^T) = 0$. In order to avoid such a nonlinear constraint we observe (see [25]) that the equation (4) is satisfied by every matrix $P + \rho I$. Therefore, we search for an arbitrary solution Q of (4) and then we search for a value for ρ so that $\det(P + P^T) = 0$ for $P = Q - \rho I$. We replace P in (4) with Q and rewrite the equation using the components of the vectors as follows

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} x & y & 1 & 0 & 0 & 0 & -x^2 & -xy & -x \\ 0 & 0 & 0 & x & y & 1 & -xy & -y^2 & -y \end{pmatrix} \boldsymbol{Q} \tag{6}$$

with

$$\boldsymbol{Q} = \begin{pmatrix} Q_{11} & Q_{12} & Q_{13} & Q_{21} & Q_{22} & Q_{23} & Q_{31} & Q_{32} & Q_{33} \end{pmatrix}^T$$

If m motion field vectors are used we have to invert a $(2m \times 9)$ matrix. It is easy to observe that the ninth column of such a matrix will be the linear combination of the first and the fifth column. Thus, the null space of the matrix contains the vector $(1,0,0,0,1,0,0,0,1)^T$. By rewriting this vector into a matrix we obtain the identity matrix. This agrees with our observation that the addition of a multiple of the identity gives a further solution to (4). To obtain a solution from (6) we set $Q_{33} = 0$. The corresponding value for ρ is $-P_{33} = -v_z N_z$. The solution for $\boldsymbol{v}$ and $\boldsymbol{N}$ can then be computed from the eigensystem of $Q + Q^T$.

We will not conduct an error analysis for this particular solution technique. We are rather interested in a method-independent error analysis technique which is provided by the Cramer-Rao theory. However the above description clarifies the way we will choose the intermediate parameters used in the next steps.

3 Cramer-Rao inequality

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 [26]

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

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 [26], the error covariance of an *unbiased* estimator is bounded below by the inverse of the Fisher information matrix:

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

An unbiased estimator that achieves the above lower bound is called *efficient*. 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 equal 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 with the estimate as the center

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

The probability that the true value $\boldsymbol{p}$ lies inside the ellipsoid is given by the constant c which geometrically expresses the ellipsoid's stretching. The directions of the symmetry axes of the ellipsoid are given by the eigenvectors of F. The lengths of the semiaxes 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 an oxymoron, 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 problem 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. We will discuss this fact in the next section in order to elucidate the coupling between translation and rotation.

4 Computation of the Fisher information matrix

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 $\sigma^2 I$. The assumptions of isotropy and constancy of the measurement noise do not hold for the optical flow measurements. Uncertainty in optical flow estimation is well known to depend on the richness of the gray-value structure. Modeling this uncertainty in order to incorporate it into our computation is a future direction of research. Under the above assumptions the conditional probability density function reads as follows

$$p(\mathcal{Z}|\boldsymbol{p}) = \frac{1}{k}\exp(-\frac{1}{2\sigma^2}\iint_D \|\dot{\tilde{\boldsymbol{x}}} - \boldsymbol{h}(\boldsymbol{p})\|^2\,dx\,dy), \qquad (10)$$

where $\boldsymbol{h}(\boldsymbol{p})$ is the measurement function we will describe below. The constant k is chosen appropriately to normalize the probability density function. 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 ego-motion of the camera.

Differentiation with respect to $\boldsymbol{p}$ yields

$$\frac{\partial \ln p(\mathcal{Z}|\boldsymbol{p})}{\partial \boldsymbol{p}} = \frac{1}{\sigma^2}\iint_D (\dot{\tilde{\boldsymbol{x}}} - \boldsymbol{h}(\boldsymbol{p}))^T \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{p}}\,dx\,dy \qquad (11)$$

and

$$\begin{aligned} F &= E[\frac{\partial \ln p(\mathcal{Z}|\boldsymbol{p})}{\partial \boldsymbol{p}}^T \frac{\partial \ln p(\mathcal{Z}|\boldsymbol{p})}{\partial \boldsymbol{p}}] \\ &= \frac{1}{\sigma^2}\iint_D \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{p}}^T \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{p}}\,dx\,dy. \end{aligned}$$

The inverse of the Fisher information matrix is proportional to the variance of the measurement noise as expected. For the sake of simplicity we will omit σ^2 in the further computations.

We return to the measurement function of the motion field (2) and collect the terms including unknowns in an intermediate parameter vector $\boldsymbol{q}$ and obtain

$$\begin{aligned} \dot{\tilde{\boldsymbol{x}}} &= B\boldsymbol{q} = \begin{pmatrix} 1 & x & y & 0 & 0 & 0 & x^2 & xy \\ 0 & 0 & 0 & 1 & x & y & xy & y^2 \end{pmatrix} \boldsymbol{q} \quad \text{with} \\ \boldsymbol{q} &= (v_x N_z + \omega_y,\ v_x N_x - v_z N_z,\ v_x N_y - \omega_z, \\ &\quad v_y N_z - \omega_x,\ v_y N_x + \omega_z,\ v_y N_y - v_z N_z, \\ &\quad \omega_y - v_z N_x,\ -v_z N_y - \omega_x)^T. \end{aligned}$$

We used a different symbol $\dot{\tilde{\boldsymbol{x}}}$ for the motion field vector in $\mathbb{R}^2$ in contrast to $\dot{\boldsymbol{x}}$ in (2) which belongs to $\mathbb{R}^3$ with the third component equal to zero. The elements of the vector $\boldsymbol{q}$ correspond to the elements of Q in (6) after rearranging columns, inserting $\rho = -v_z N_z$ and negating Q_{31} and Q_{32}.

Hence, the derivative of the measurements function with respect to the unknown parameters $\boldsymbol{p}$ may be written

$$\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{p}} = \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{q}}\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{p}} = B\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{p}}.$$

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

$$F = \frac{\partial \boldsymbol{q}}{\partial \boldsymbol{p}}^T \left\{\iint_D B^T B\,dx\,dy\right\} \frac{\partial \boldsymbol{q}}{\partial \boldsymbol{p}}. \qquad (12)$$

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

$$B_{integral} = \iint_D B^T B\,dx\,dy$$

depends only on the size of the field of view and its determinant reads

$$\det(B_{integral}) = \frac{1}{25}\alpha\beta \mathrm{A}^3\mathrm{B}^3(4\mathrm{A}+5\mathrm{B})(5\mathrm{A}+4\mathrm{B}). \qquad (13)$$

Thus, the error covariance is a monotonically decreasing function of the size of the effective field of view.

We hence omit the factor $\alpha\beta$, too, in order to simplify the further expressions.

Before we proceed with the computation of the Jacobian $\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{p}}$ 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:

$$\begin{aligned} N'_x &= \tfrac{N_x}{N_z} & v'_x &= v_x N_z \\ N'_y &= \tfrac{N_y}{N_z} & v'_y &= v_y N_z \\ & & v'_z &= v_z N_z. \end{aligned} \qquad (14)$$

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:

$$\boldsymbol{p} = \begin{pmatrix} v_x & v_y & v_z & \omega_x & \omega_y & \omega_z & N_x & N_y \end{pmatrix}.$$

We obtain the following Jacobian for the measurement function

$$\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{p}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ N_x & 0 & -1 & 0 & 0 & 0 & v_x & 0 \\ N_y & 0 & 0 & 0 & 0 & -1 & 0 & v_x \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & N_x & 0 & 0 & 0 & 1 & v_y & 0 \\ 0 & N_y & -1 & 0 & 0 & 0 & 0 & v_y \\ 0 & 0 & -N_x & 0 & 1 & 0 & -v_z & 0 \\ 0 & 0 & -N_y & -1 & 0 & 0 & 0 & -v_z \end{pmatrix}. \qquad (15)$$

Before we recover its inverse we compute its determinant:

$$\det(F) = \det{}^2(\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{p}})\det(B_{integral})$$

After tedious adding and subtracting the rows of the Jacobian we obtain

$$\det(\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{p}}) = \|\boldsymbol{N} \times \boldsymbol{v}\|^2. \qquad (16)$$

This result is new in such a general form – only the third element of the normal $\boldsymbol{N}$ is assumed to be equal unity –

although it has been already stated in the ambiguity framework [18, 7, 32]: The case of parallel normal and translation causes the existence of a unique solution, but as a degenerate case of the general one of two solutions this unique solution is sensitive to noise. We observe that in this degenerate case the Fisher information matrix becomes singular and the Cramer-Rao lower bounds are infinitely large.

Furthermore, the Fisher information matrix is independent of the angular velocity. This observation is trivial as already argued by [37] since the measurement function is linear in $\boldsymbol{\omega}$.

We carry out the matrix multiplications in (13) and obtain

$$F = \begin{pmatrix} K & L \\ L^T & M \end{pmatrix}. \tag{17}$$

The reader is referred to [?] for the long expressions of the submatrices omitted here due to space limitation.

5 Confusion between translation and rotation

The motion field vector is the sum of two components, a translational one including the information about the environment and a rotational one (see also eq. (2)). 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).

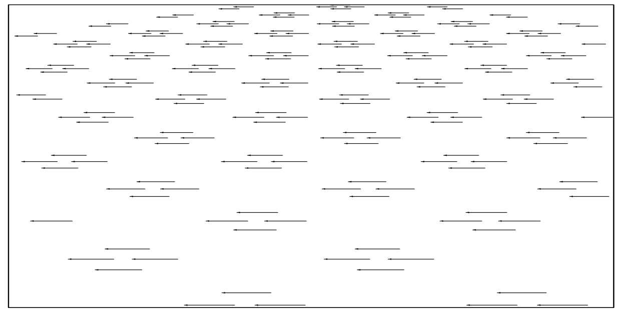
This confounding becomes dominant if the field of view is small or if there is no depth variation in the environment. However, as already argued by other authors, one may robustly compute the amount of motion represented by the sum $v_x N_z + \omega_y$ and the difference $v_y N_z - \omega_x$ since they build the zeroth order terms regarding the motion field as a polynomial with respect to the image coordinates (x, y). In the following we will show by means of the lower bound of the error covariance that such "robust" combinations of unknowns do indeed exist although the estimate for each individual unknown is sensitive to noise.

In order to invert the Fisher information matrix computed in the last section we make use of the formula [8]

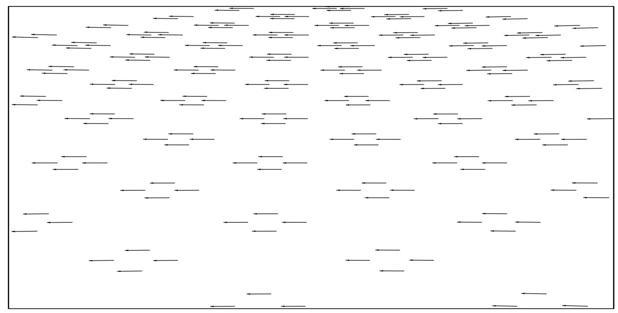
$$\begin{pmatrix} K & L \\ L^T & M \end{pmatrix}^{-1} = \begin{pmatrix} E^{-1} & -E^{-1}LM^{-1} \\ -M^{-1}L^TE^{-1} & M^{-1} + M^{-1}L^TE^{-1}LM^{-1} \end{pmatrix}$$
$$\text{with} \quad E = (K - LM^{-1}L^T) \tag{18}$$

The Fisher information matrix is a function of the field of view and the magnitude of the scaled translational velocity as well as a function of the directions of the translational velocity $\boldsymbol{v}$ and the normal $\boldsymbol{N}$. The matrix E obtains the following block-diagonal form

$$E = \begin{pmatrix} E_{135} & 0 \\ 0 & E_{246} \end{pmatrix}, \tag{19}$$



(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 motion fields have been produced by simulation of the motion of a camera on a gripper in front of a calibration plate.

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 as illustrated in Fig. 2. We introduce the angles ψ and χ between the optical axis and $\boldsymbol{v}$ and $\boldsymbol{N}$, respectively:

$$\begin{aligned} \boldsymbol{v} &= \begin{pmatrix} v_x & 0 & v_z \end{pmatrix} = \begin{pmatrix} \|\boldsymbol{v}\|\sin\psi & 0 & \|\boldsymbol{v}\|\cos\psi \end{pmatrix} \\ \boldsymbol{N} &= \begin{pmatrix} N_x & 0 & 1 \end{pmatrix} = \begin{pmatrix} \tan\chi & 0 & 1 \end{pmatrix}. \end{aligned}$$

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.

$$E^{-1} = \begin{pmatrix} E_{135}^{-1} & 0 \\ 0 & E_{246}^{-1} \end{pmatrix}. \tag{20}$$

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

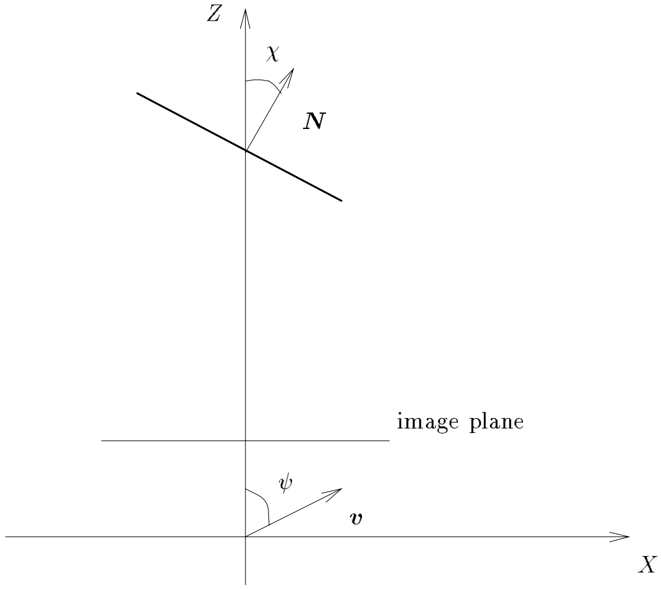


Figure 2: Illustration of the used angles in case of coplanar viewing direction, plane normal and translational velocity ($N_y = 0, v_y = 0$). We denote the angle between viewing and translation direction by ψ and the angle between viewing direction and the normal by χ.

analogously. The diagonal elements corresponding to the lower bounds of the variances of v_x, v_z and ω_y are

$$
\begin{aligned}
(E_{135}^{-1})_{11} &= \frac{10\mathrm{A} + 8\mathrm{A}\cos^2\psi + (14\mathrm{A}^2+5)\sin^2\psi + 9\mathrm{A}\tan\chi\sin\psi(\cos\psi - \tan\chi\sin\psi)}{9\mathrm{A}^2(\tan\chi\cos\psi - \sin\psi)^2} \\
(E_{135}^{-1})_{22} &= \frac{1}{\mathrm{A}} \\
(E_{135}^{-1})_{33} &= \frac{5\sin^2\psi + 18\mathrm{A}\cos^2\psi + 28\mathrm{A}\tan\chi\sin\psi\cos\psi + \mathrm{A}\tan^2\chi(14\mathrm{A}\cos^2\psi + 9\sin^2\psi)}{9\mathrm{A}^2(\tan\chi\cos\psi - \sin\psi)^2}
\end{aligned}
$$

The singularity induced by the terms in the denominators expresses the case $\chi = \psi$ of parallel translation and normal already proved by the computation of the determinant of the Fisher information matrix.

We note that the uncertainty in the estimate for v_z is independent of the translation and the normal and it is not affected by the singularity $\boldsymbol{v} \parallel \boldsymbol{N}$.

We next focus on the uncertainty in the parameter space (v_x, ω_y). 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 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 [9]:

$$\lambda_{min}(S) \leq \boldsymbol{u}^T E_{135}^{-1}\boldsymbol{u} \leq \lambda_{max}(S). \tag{21}$$

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 for two sizes of the field of view. Fig. 3 shows that the smallest eigenvalue is not affected by the singularity $\chi = \psi$. However, the error variances of v_x and ω_y in (22) become infinitely large. This fact substantiates our methodology in exploiting the entire structure of the lower bound covariance matrix. Fig. 3 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} (see Fig. 4) with help of MAPLE and plot it in the same way as above. Fig. 5 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 $v_x + \omega_y$ 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 fields 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. 5 below). The angle ϕ_{min} may take values greater than $\pi/4$ but is never close to $\pi/2$ what prevents the estimate for ω_y from having the lowest uncertainty.

The analysis in (v_y, ω_x)-space can be carried out in the same way. We found out that the direction of lowest uncertainty in case of a small field of view is $-\pi/4$ which allows a robust computation of the difference $-v_y + \omega_x$.

6 Uncertainty in the computation of plane's normal

The information about the uncertainty lower bounds in the direction (N_x, N_y) of the normal is contained in the lower-right submatrix of the inverse of the Fisher information matrix in (19). We denote this submatrix by D:

$$D = M^{-1} + M^{-1}L^T E^{-1} L M^{-1}. \tag{22}$$

After applying the same assumptions $v_y = 0, N_y = 0$ for the normal and the translation, D becomes diagonal and we obtain the following variances for N_x and N_y:

$$D_{11} = \frac{18\mathrm{A} + (5 + 28\mathrm{A} + 14\mathrm{A}^2)\tan^2\chi + 9\mathrm{A}\tan^4\chi}{9\mathrm{A}^2\|\boldsymbol{v}\|^2(\tan\chi\cos\psi - \sin\psi)} \quad (23)$$

$$D_{22} = \frac{18\mathrm{A} + (5 + 28\mathrm{A} + 14\mathrm{A}^2)\tan^2\chi}{9\mathrm{A}^2\|\boldsymbol{v}\|^2(\tan\chi\cos\psi - \sin\psi)} \quad (24)$$

Both are singular if the translation is parallel to the normal. The lower bounds grow if the scaled translation magnitude becomes smaller. This has been expected since vanishing translation does not allow the recovery of depths. In the function plots of the variances of N_x and N_y (Fig. 6) we use the arctan of the variance in order to include the case of infinite values. As an artifact of our modeling we obtain an infinite variance for planes parallel to the optical axis, too, since in this case the visible part of the plane inside the field of view corresponds to infinite depths.

We observe that the estimates are less sensitive if the plane is frontal ($\chi = 0$) and the translation is parallel to the image plane ($\psi = \pi/2$). Hence, we get a trade-off between motion and structure computation: A geometry-motion configuration that enables a robust estimation of the normal (small values for the arctan of the error variance – see Fig. 6) causes a sensitive estimation for motion, i.e. significantly non-zero values for the smallest eigenvalue of S – see Fig. 3.

7 Conclusion

We have 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 uncertainty lower bounds are due to Cramer-Rao and are valid for any unbiased estimator under the assumption of Gaussian zero-mean noise in the motion field. In order to invert the Fisher information matrix and to reduce the number of the parameters affecting the sensitivity, we have restricted our analysis to the case of coplanar translational velocity, normal and viewing direction. The error variance becomes infinite if the translation is parallel to the normal (17). Moreover, the error variance is a decreasing function of the size of the field of view – see the factor $\alpha\beta$ in (**??**).

The parameters of the normal can be estimated more robustly if the translation is parallel to the image plane and the plane's slant is small. We, thus, show a trade-off between structure and motion estimation regarding sensitivity.

Next research steps include the sensitivity analysis of motion estimation from multiple frames (see also [23, 29]). We developed a recursive algorithm and tested it on real as well as synthetic experiments[1] results show that if the motion is purely translational the coupling between translation and rotation persists along time.

Future work in error sensitivity has to be done for the new active and/or qualitative motion estimation techniques in order to provide rigorous stability proofs that will substantiate the successful real-world experiments.

[1] Citation omitted in order not to easily jeopardize the double blind review process.

Acknowledgements

The financial support of the first author by the German Academic Exchange Service (DAAD) is gratefully acknowledged. We thank Y. Aloimonos, A. Zisserman, D. Murray, F. Bergholm, H. Sawhney and V. Sundareswaran for fruitful and motivating discussions.

References

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

[2] J.K. Aggarwal and N. Nandhakumar. On the computation of motion from sequences of images - a review. *Proceedings of the IEEE*, 76:917–935, 1988.

[3] 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.

[4] 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.

[5] B.F. Buxton, D.W. Murray, H. Buxton, and N.S. Williams. Structure from motion algorithms for computer vision on an simd architecture. *Computer Physics Communications*, 37:273–280, 1985.

[6] 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.

[7] O.D. Faugeras and F. Lustman. Motion and structure from motion in a piecewise planar environment. *International Journal of Pattern Recognition and Artificial Intelligence*, 2:485–508, 1988.

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

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

[10] J.C. Hay. Optical motions and space perception: an extension of gibson's analysis. *Psychological Review*, 73:550–565, 1966.

[11] 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.

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

[13] B.K.P. Horn and E.J. Weldon. Computationally-efficient methods for recovering translational motion. In *Proc. Int. Conf. on Computer Vision*, pages 2–11, London, UK, June 8-11, 1987.

[14] C.P. Jerian and R. Jain. Structure from motion - a critical analysis of methods. *IEEE Trans. Systems, Man, and Cybernetics*, 21:572–587, 1991.

[15] K. Kanatani. Structure and motion from optical flow under perspective projection. *Computer Vision, Graphics, and Image Processing*, 38:122–146, 1987.

[16] 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.

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

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

[19] H.C. Longuet-Higgins. The reconstruction of a plane surface from two perspective projections. *Proc. Royal Society of London*, B227:399–410, 1986.

[20] S.J. Maybank. The angular velocity associated with the optical flow field arising from motion through a rigid environment. *Proc. Royal Society of London*, A401:317–326, 1985.

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

[22] D.W. Murray, D.A. Castelow, and B.F. Buxton. From image sequences to recognized moving polyhedral objects. *International Journal of Computer Vision*, 3:181–208, 1989.

[23] D.W. Murray and D.M. Pickup. Recursive updating of planar motion. In *Proc. British Machine Vision Conference*, pages 169–177, Glasgow, UK, Sept. 24-26, 1991.

[24] S. Negahdaripour. Closed-form relationship between the two interpretations of a moving plane. *Journal Opt. Soc. Am.*, A7:279–285, 1990.

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

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

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

[28] M. Subbarao and A.M. Waxman. Closed form solutions to image flow equations for planar surfaces in motion. *Computer Vision, Graphics, and Image Processing*, 36:208–228, 1986.

[29] S. Sull and N. Ahuja. Estimation of motion and structure of planar surfaces from a sequence of monocular images. In *IEEE Conf. Computer Vision and Pattern Recognition*, pages 732–733, Maui, Hawaii, June 3-6, 1991.

[30] G. Toscani and O.D. Faugeras. Structure and motion from two perspective views. In *Proc. IEEE Int. Conf. on Robotics and Automation*, pages 221–227, Raleigh, North Carolina, March 31 - Apr. 2, 1987.

[31] 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*, PAMI-6:13–27, 1984.

[32] R.Y. Tsai, T.S. Huang, and W. Zhu. Estimating three-dimensional motion parameters of a rigid planar patch, ii: Singular value decomposition. *IEEE Trans. Acoustics, Speech, and Signal Processing*, ASSP-30:525–534, 1982.

[33] A. Verri, F. Girosi, and V. Torre. Mathematical properties of the 2d motion field: from singular points to motion parameters. *Journal Opt. Soc. Am.*, A6:698–712, 1989.

[34] 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.

[35] J. Weng, N. Ahuja, and T.S. Huang. Motion and structure from point correspondences with error estimation: planar surfaces. *IEEE Trans. Signal Processing*, 39:2691–2717, 1991.

[36] J. Weng, T.S. Huang, and N. Ahuja. Motion and structure from two perspective views: algorithms, error analysis, and error estimation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, PAMI-11:451–476, 1989.

[37] 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*, PAMI-14:995–1013, 1992.

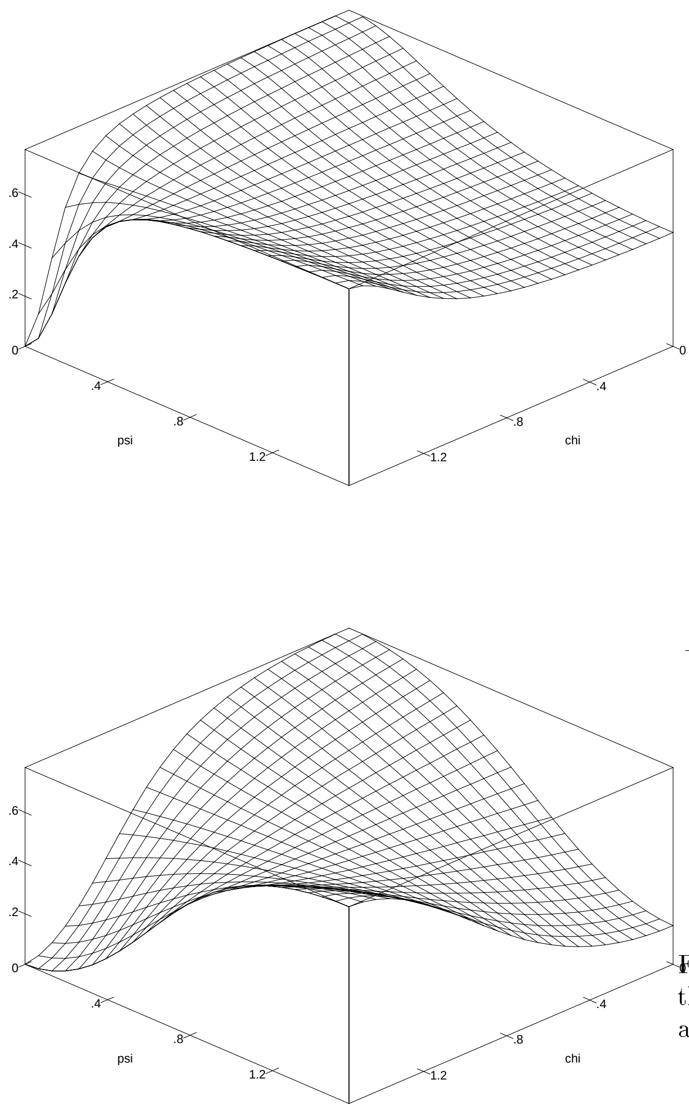


Figure 3: The smallest eigenvalue of S as a function of the angles ψ of translation and χ of the normal with the optical axis for two sizes of the field of view: A = 0.1 (above) and A = 1.0 (below).

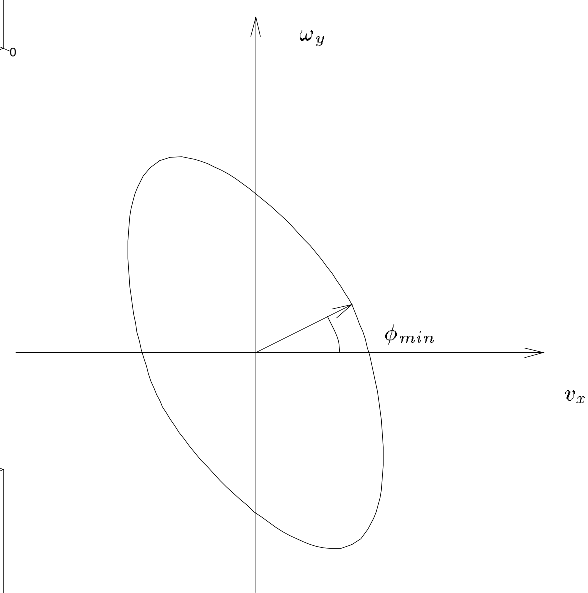
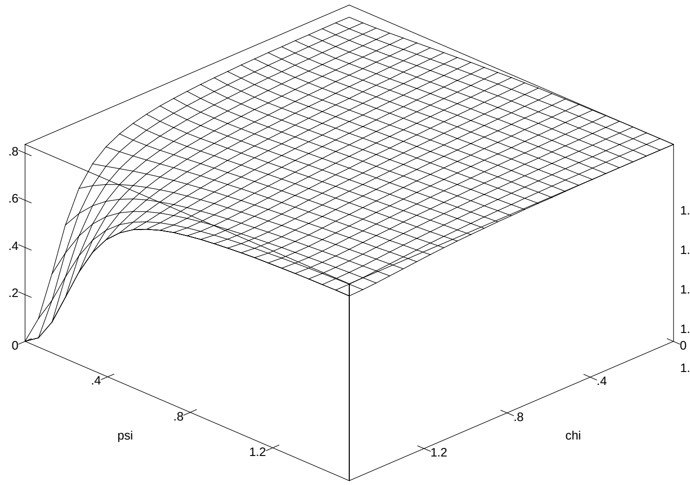
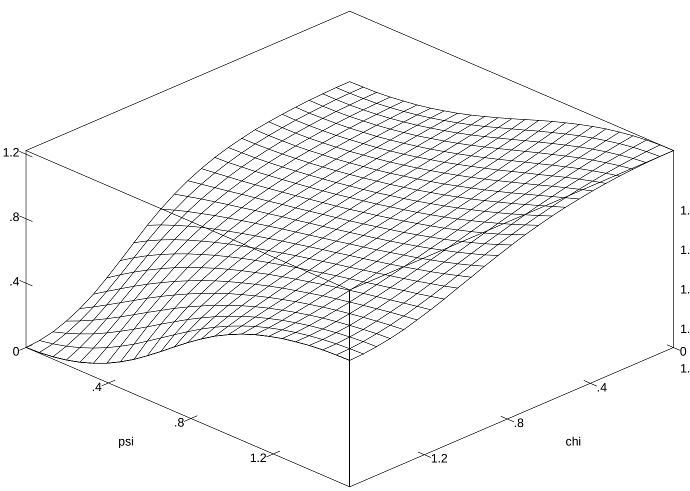


Figure 4: The intersection of the error ellipsoid 9 with the plane (v_x, ω_y) yields the uncertainty ellipse. The angle ϕ_{min} gives the direction of lowest uncertainty.



(a)



(b)

Figure 5: 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.

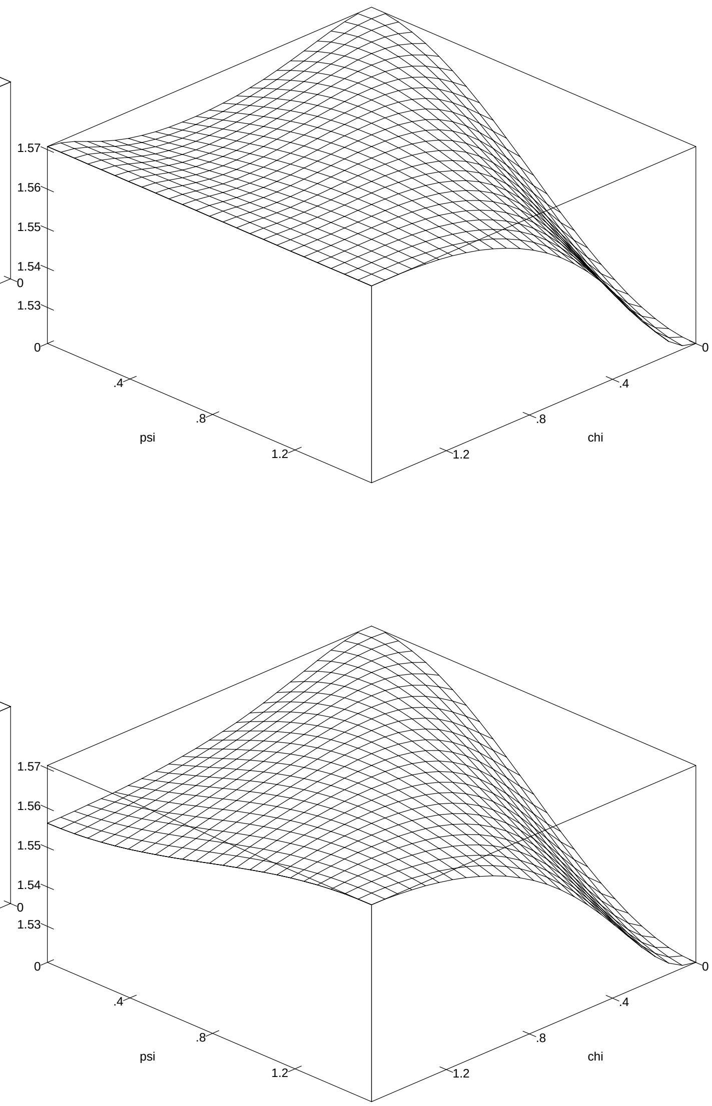


Figure 6: The arctan of the error variance of N_x (above) and N_y (below) as a function of the angles ψ of translation and χ of the normal with the optical axis for size A = 0.1 of the field of view. The value $\pi/2$ represents an infinite error variance and appears in case of parallel translation and normal ($\chi = \psi$).