A "Half-Perspective" Approach to Robust Ego–Motion Estimation for Calibrated Cameras *

Robert Wagner
Universität Passau, Fakultät für Mathematik und Informatik,
Innstraße 29, D-94032 Passau, Germany,
e-mail: wagner@fmi.uni-passau.de.

Klaus Donner
Universität Passau, Fakultät für Mathematik und Informatik,
Innstraße 29, D-94032 Passau, Germany,
e-mail: donner@fmi.uni-passau.de.

Feiyu Liu
BMW AG, FIZ, Abt. EW-1,
Postfach 400240, D-80788 München, Germany,
e-mail: feiyu@bmwmun.bmw.de.

December 4, 1997

*This work was supported by the "Bayerische Motorenwerke Aktiengesellschaft" and by the German Ministry for Research and Education (BMBF) under Grant No. 01M3010A5.

Abstract

A new computational approach to estimate the ego-motion of a camera from sets of point correspondences taken from a monocular image sequence is presented. The underlying theory is based on a decomposition of the complete set of model parameters into suitable subsets to be optimized separately, e.g. all stationary parameters concerning camera calibration are adjusted in advance (calibrated case).

The first part of the paper is devoted to the description of the mathematical model, the so-called conic error model, and the numerical solution of the derived optimization problem. In contrast to existing methods, the conic error model permits to distinguish between feasible and non-feasible image correspondences related to 3D object points in front of and behind the camera, respectively. Based on this "half-perspective" point of view, a well-balanced objective function is derived that encourages the proper detection of mismatches and distinct relative motions.

In the second part, the results of various tests are presented and analyzed. The experimental study clearly shows that the numerical stability of the new approach is superior to that of so-called self-calibration techniques (uncalibrated case). Furthermore, the precision of the estimates is better than that achieved by comparable methods in the calibrated case based on a "full-perspective" modeling and the related epipolar geometry. Accordingly, the accuracy of the resulting ego-motion estimation turns out to be excellent, even without any further temporal filtering.

Keywords

Computer vision, motion estimation, perspective geometry, two-view analysis, image sequence analysis.

1 Introduction

For many years, the computation of the 3D ego-motion of a camera from pairs of corresponding geometric primitives has been of considerable interest in computer vision.

Based on a full perspective projection model, Roach and Aggarwal [1] proved in the early eighties that two views are sufficient to recover the camera motion from point correspondences. Since then, many linear and non-linear methods have been developed for calibrated as well as uncalibrated cameras.

Using eight pairs of corresponding points from two subsequent images, Tsai and Huang [2] and Longuet-Higgins [3] derived solutions based on eight linear equations. These methods have been improved by many researchers to achieve better accuracy with noisy data (e.g. Spetsakis and Aloimonos [4], Weng, Huang and Ahuja [5, 6] or Philip [7]). A comprehensive review is given in [8]. For a detailed mathematical background the reader is referred to [9].

Recently, several authors in computer vision and robotics have investigated so-called self-calibration techniques featuring online camera calibration. In [10, 11] Luong and Faugeras introduced a 3×3 matrix called the *Fundamental Matrix*, which reduces in the calibrated case to the *Essential Matrix* proposed in the original Longuet-Higgins equations [3]. The elements of the *Fundamental Matrix* are composed of the set of intrinsic camera model parameters corresponding to an ideal central projection and the motion parameters. This technique has been used by Faugeras and Robert [12], Viéville, Faugeras and Luong [13] to recover camera motion in the uncalibrated case.

The reconstruction of the location and motion of objects from a sequence of perspective images taken by a moving camera is known to be a highly instable inverse problem. The stability and precision of the achieved results is quite sensitive to the accuracy of the assumed camera model, its intrinsic parameters, the results of ego-motion estimation and the extracted image correspondences. Our studies of various approaches show that the decomposition of the global (camera calibration and motion estimation) optimization task into sub-tasks – and to solve each of those separately – avoids a compensation of errors related to distinct model parameters (which thus yielded formal error minimization within the global optimization procedure but still inaccurate model parameters).

In addition to general advantages, such as improved stability and precision of the relevant optimization procedures and the reduction of the overall computation time, a decomposition encourages the

- simplification and generalization of sub-models: the mathematical transformation describing the imaging of the camera, for instance, can easily be extended to include non-linear distortions from the ideal central projection as discussed in [14];
- integration of certain a-priori knowledge about the system (e.g. restricted ranges of motion parameters such as small rotational angles or planar movement of the camera) to relieve the optimization processes of unnecessary computational effort;
- detailed study of numerical characteristics of the individual optimization tasks and, therefore, the design of fast and numerically stable algorithms.

The overall quality and accuracy of the motion estimation depend particularly on the correct and sufficiently precise modeling and treatment of all error sources. Both random and systematic distortions – such as digitization noise, projective distortions, poorly matched or even mismatched pairs of geometric tokens and correspondences related to object points with a different relative 3D motion with respect to the camera – have to be taken into account.

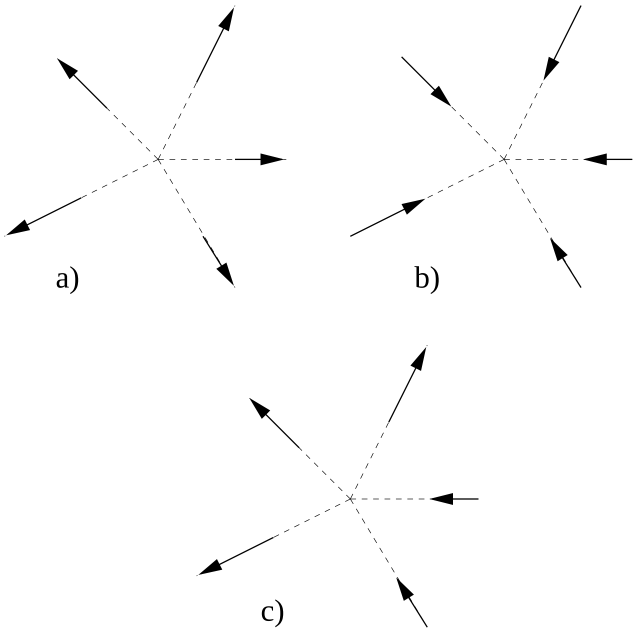

Figure 1: “Ambiguous” examples of image point correspondences considering a pure translation of the camera

Fig. 1 shows three examples of image point correspondences. The arrows indicate the displacement in the image plane for the projections of 3D object points before and after a motion of the camera. Considering a pure translation of the image sensor, Fig. 1.a) depicts sample correspondences for a forward movement of the camera (i.e. towards the related 3D object points). In this case, the image points move away from the so-called *epipole* or *focus of expansion* (i.e. the projection of the translational direction). Fig. 1.b) depicts the inverse situation, a backward movement of the camera and an induced movement of the image points towards the epipole. In Fig. 1.c), however, some of the image points move towards the epipole, and others away. This implied that the camera is moving forward and backward (or that 3D object points are located in front and behind the camera) at the same time, which is impossible. Hence, (provided that the epipole has been determined correctly) this constellation reflects an example in which mismatches or various 3D objects with different relative motions occur.

From a full-perspective point of view and an interpretation based on the epipolar geometry, where no distinction is made between object points in front of and behind the camera, all three depicted constellations are “legal” and treated identically. In this case, only the distance of the corresponding image point after the motion of the camera from the epipolar line (the line through the epipole and the image point before motion, visualized in dashed lines) is taken into account. This distance is zero for all depicted correspondences in Fig. 1. Therefore, only an integration of the direction of the displacement for pairs of corresponding image points by a restriction of the underlying model to 3D object points located in front of the camera (half perspective model) permits a robust detection of such "ambiguous" correspondences. Consequently, the sign of the translational movement can be uniquely determined from the set of given image correspondences.

In this paper, the special case of point correspondences as pairs of tokens is considered. However, the main ideas presented in this and the subsequent sections can also be applied to design algorithms that use more complex geometric primitives, e.g. line segments.

The paper is organized as follows: In section 2, an abstract camera model is introduced and a suitable error weight function (i.e. perspectively invariant and restricted to object points in front of the camera) is developed based on a conic model. The subsequent section 3 is devoted to the statement of the problem of motion estimation. Furthermore, it is shown how to interpret point correspondences in the calibrated case by means of viewing rays and cones in 3D, and how to apply the conic error model to derive the general non-linear objective function. The last part of the section deals with details of the implementation, potential simplifications of the model and the generation of suitable initial values for the non-linear optimization procedure. Section 4 provides a variety of experimental results for both synthetic and natural data, investigating

- the numerical stability in relation to the 3D distance and image position of objects,
- the precision of the estimated motion parameters (including a comparison with gyro-senor data and results based on a full-perspective model proposed by Weng, Huang and Ahuja in [5, 6]),
- the stability of the derived results compared to those of self-calibration approaches and
- the effect of a series of special approximations suitable for automotive applications and the generation of initial values for the non-linear optimization procedures.

A last section concludes with a discussion of the results presented in section 4.

2 Camera Model and Error Approximation

The camera model introduced in this section is described by a generalized point to point mapping between the Euclidean 3D space and the 2D image plane. Perspective transformations are treated as a special case. Furthermore, the so-called conic error model is presented to achieve the mathematical description of the various error contributions for the generalized camera model.

2.1 Notation

Throughout this paper, small letters $a, b, c, \ldots$ and capital letters $A, B, C, \ldots$ denote column vectors and matrices, respectively. Superscripts ^{T} indicate the transpose of a vector or a matrix. The inner product of two vectors a and b is denoted by $\langle a|b\rangle$ and their cross-product by $a \times b$. $\|a\| = \sqrt{\langle a|a\rangle}$ represents the Euclidean norm of the vector a, $S_1(0) = \{x \in I\!R^3 \mid \|x\| = 1\}$ the unit sphere in $I\!R^3$ and I_3 the unit matrix in $I\!R^3$.

The absolute angle between a plane in 3D with normal vector n and distance δ from the origin,

$$E = \{x \in I\!R^3 \mid \langle n|x\rangle = \delta\},$$

and a semi-line starting from a point $c \in E$ with direction vector d,

$$r = \{x \in I\!R^3 \mid x = c + \lambda d, \lambda \in I\!R, \lambda \geq 0\},$$

is given by

$$\sphericalangle\,(r, E) = \arcsin \frac{|\langle n|d\rangle|}{\|n\|\;\|d\|} \in [0; \frac{\pi}{2}].$$

The angle between two semi-lines

$$\begin{aligned} r_1 &= \{x \in I\!R^3 \mid x = c + \lambda d_1, \lambda \in I\!R, \lambda \geq 0\}, \\ r_2 &= \{x \in I\!R^3 \mid x = c + \lambda d_2, \lambda \in I\!R, \lambda \geq 0\}, \end{aligned}$$

and/or their direction vectors d_1 and d_2 is written as

$$\sphericalangle\,(r_1, r_2) = \sphericalangle\,(r_1, d_2) = \sphericalangle\,(d_1, r_2) = \sphericalangle\,(d_1, d_2) = \arccos \frac{\langle d_1|d_2\rangle}{\|d_1\|\;\|d_2\|} \in [0; \pi].$$

2.2 Camera Model

The purpose of camera modeling and camera calibration is to describe and to determine the mapping $\Pi : I\!R^3 \to I\!R^2$ between points in a given 3D coordinate system and their pixel positions in the image plane.

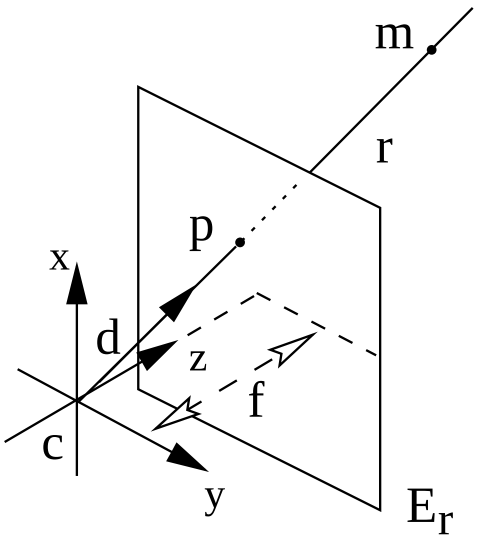

Figure 2: Pinhole camera model and perspective projection

For a pinhole camera (Fig. 2), for instance, the mapping is given by an ideal perspective projection with image plane E_r and projection center $c \notin E_r$. The camera coordinate system is a 3D Cartesian system with origin c. By convention, the z-axis is perpendicular while the x- and y-axes are parallel to the image plane E_r. The distance f of the projection center c from the image plane E_r is referred to as *focal length*. The projection p of a point $m = [m_x, m_y, m_z]^T$ satisfying $m_z \neq 0$ is defined as the intersection of the straight line $\overline{cm}$ passing through c and m with the image plane E_r. Accordingly, Π can be written as

$$p = \Pi(m) = \left[f\frac{m_x}{m_z}, f\frac{m_y}{m_z} \right]^T ;$$

m_x, m_y and m_z denote the 3D-coordinates of the point m in the camera coordinate system.

This model, however, may not be accurate enough for real optical systems. Therefore, many authors have introduced more complex camera models and calibration procedures in order to describe the qualitative and quantitative physical properties of the camera in more detail. In [14], Tsai proposed a four step model featuring a decomposition of the mapping between points in a given 3D coordinate system and their pixel positions in the image plane into

1. a transformation from a given 3D coordinate system to the 3D camera coordinate system, described by a rotation R and a translation t,

2. the perspective projection onto the image plane of an ideal pinhole camera,

3. a subsequent correction of the radial and tangential distortions caused by the lens system and, finally,

4. a compensation of the effects resulting from scanning and sampling the image.

The transformation mentioned in (1.) includes the so-called *extrinsic* camera parameters and determines the location and orientation of the image sensor in a given 3D coordinate system. The mappings associated with (2.), (3.) and (4.) depend on the *intrinsic* camera parameters. Both, intrinsic as well as extrinsic parameters are estimated by the camera calibration process. In the calibrated case, the calibration is usually achieved interactively, taking advantage of some suitable 3D geometric pattern and error minimization procedure.

In general, Π can be interpreted in an abstract sense as composition $\Pi_3 \circ \Pi_2 \circ \Pi_1$ of a transformation between two 3D Euclidean coordinate systems, $\Pi_1 : I\!R^3 \to I\!R^3$ (step 1), an ideal perspective projection $\Pi_2 : I\!R^3 \to I\!R^2$ (step 2) and a potentially non-linear mapping $\Pi_3 : I\!R^2 \to I\!R^2$ (steps 3 and 4) that converts pixel positions of the ideal pinhole camera to pixel coordinates of the physical optical system.

If Π_3 is an affine transformation, the composition $\Pi = \Pi_3 \circ \Pi_2 \circ \Pi_1$ is projectively linear, i.e. using homogeneous coordinates, Π is a linear transformation and can be written in matrix form.

Example 1 *Consider a simple pinhole camera with*

- *focal length f,*

- *a horizontal size of pixels s_u,*

- *a vertical size of pixels s_v,*

- *pixel coordinates (u_0, v_0) of the center of optics (i.e. the intersection between the optical axis and the image plane), and*
- *an Euclidean displacement (a rotation R and a translation t) between the world and camera coordinate system.*

Then, using homogeneous coordinates, the mappings Π_1, Π_2 and Π_3 can be written as

$$
\begin{aligned}
\Pi_1(m) &= Rm + t = A_{\Pi_1}[m, 1]^T, \\
[\Pi_2(m), 1]^T &= \left[f\frac{m_x}{m_z}, f\frac{m_y}{m_z}, 1\right]^T = [fm_x, fm_y, m_z]^T = A_{\Pi_2} m, \\
[\Pi_3(u, v), 1]^T &= \left[\frac{u}{s_u} + u_0, \frac{v}{s_v} + v_0, 1\right]^T = A_{\Pi_3}[u, v, 1]^T, \\
[\Pi(m), 1]^T &= A_\Pi [m, 1]^T,
\end{aligned}
$$

where the matrices A_{Π_1}, A_{Π_2}, A_{Π_3} and A_Π are defined as follows:

$$
\begin{aligned}
A_{\Pi_1} &= \begin{bmatrix} R & t \end{bmatrix}, \\
A_{\Pi_2} &= \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}, \\
A_{\Pi_3} &= \begin{bmatrix} \frac{1}{s_u} & 0 & u_0 \\ 0 & \frac{1}{s_v} & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \\
A_\Pi &= A_{\Pi_3} A_{\Pi_2} A_{\Pi_1}.
\end{aligned}
$$

Formally, the correction mapping Π_3 can be defined without explicit knowledge of intrinsic camera parameters like center of optics, pixel size, etc. Then, Π_3 can be chosen as a general 2D transformation with certain reasonable constraints: Π_3 should be one-to-one (i.e. invertible), convex, continuous and differentiable. Thus, an appropriate subset of second and higher order polynomials can be used to compensate all distortions caused by the optical lens system and the scanning/sampling process.

With the uncalibrated case based on the computation of the *Fundamental Matrix*, the mapping Π_1 is the identity transformation $\Pi_1(x) = x$, i.e. the so-called world coordinate system and the 3D camera coordinate system are identical, and Π_3 is usually assumed to be projectively linear (a possible extension of online calibration methods for cameras with lens distortion is described in [15]). Therefore, almost any self-calibration approach, despite of the advantage that interactive calibration procedures are superfluous, suffers from two significant deficiencies:

- since the mapping between the world and camera coordinate system is unknown, both, the integration of additional, non-optical sensor data (in multi-sensor applications) as well as the precise control of actuators via online image interpretation is not feasible;
- the compensation of non-linear optical distortions is not taken into account, i.e. the underlying camera model is assumed to be perspectively linear.

2.3 Modeling of Errors in the Calibrated Case

With practical applications like motion estimation and 3D reconstruction, the inverse mapping Π^{-1}, which relates image pixel positions to 3D data, plays the essential role. Due to the perspective projection Π_2, the transformation Π maps all points of a straight line

$$\{m \in \mathbb{R}^3 \mid m = c + \lambda d, \lambda \in \mathbb{R}\}$$

through the projection center c onto a distinct image point $p = \Pi(m)$.

For real optical systems, however, Π is restricted to points m located in front of the camera. According to that fact and to avoid misinterpretations of image correspondences as depicted in Fig. 1, the inverse mapping Π^{-1} is defined in terms of semi-lines by

$$r = \Pi^{-1}(p) = \{m \in \mathbb{R}^3 \mid m = c + \lambda d, \lambda \in \mathbb{R}, \lambda \geq 0\}.$$

This is readily achieved by a proper alignment of the direction vector d with the optical axis of the camera. Throughout this paper, r is called the *viewing ray* associated with the image point p. In addition to this definition, a unique representation of the viewing ray r in terms of the direction vector is obtained by assuming d to be normalized, i.e. $\|d\| = 1$.

For a given viewing ray

$$r = \{m \in \mathbb{R}^3 \mid m = c + \lambda d, \lambda \in \mathbb{R}, \lambda \geq 0\},$$

the error contributions associated with points in 3D should be weighted

- perspectively invariant: for any direction vector $d' \neq 0$ and two points

  $$\begin{aligned} m_1 &= c + \lambda_1 d', \ \lambda_1 > 0, \\ m_2 &= c + \lambda_2 d', \ \lambda_2 > 0, \end{aligned}$$

  the error $\delta(r, m_1)$ should be equal to $\delta(r, m_2)$.

- independent for all directions with the same angular difference from d: for any two normalized direction vectors d_1 and d_2 with

  $$\sphericalangle(d_1, d) = \sphericalangle(d_2, d)$$

  and two points

  $$\begin{aligned} m_1 &= c + \lambda d_1, \\ m_2 &= c + \lambda d_2, \end{aligned}$$

  the error $\delta(r, m_1)$ should be equal to $\delta(r, m_2)$.

Both requirements imply that all points with an error δ_0 from a given viewing ray r are located on a rotationally invariant cone with central vector d and origin c. Such a cone can be defined by

$$C_{c,d,\alpha} = \{x \in \mathbb{R}^3 \mid \langle d | x - c \rangle = \alpha \, \|x - c\| \}$$

with $\alpha \in [-1, 1]$, $c \in \mathbb{R}^3$ and $d \in S_1(0)$. The aperture angle of the cone $C_{c,d,\alpha}$ is $\arccos \alpha$.

Reversely, for any point $m \neq c$ and a given viewing ray r, the cone $C_{c,d,\alpha}$ containing m is unique. Furthermore, if the angular difference $\sphericalangle(d, m - c)$ between the viewing ray r and the "ideal viewing ray" starting in the projection center c and passing through the object point m is used for the error $\delta(r, m)$ of the viewing ray r with respect to the object point m, i.e.

$$\delta(r, m) = \sphericalangle(d, m - c) = \psi,$$

then $\delta(r, m)$ is exactly the aperture angle ψ of the unique cone

$$C(r, \psi) = C_{c,d,\cos(\psi)},$$

containing m. Accordingly, throughout this paper, $C(r, \psi)$ is called *error cone* for the viewing ray r with error ψ.

3 Motion Estimation from Correspondences of Image Points

3.1 Statement of the Problem

Any movement of a rigid object in 3D can be described as the composition of a rotation R and a translation t. However, the absolute value $\|t\|$ of the displacement cannot be recovered from image sequences without a-priori information about objects in the scene. Hence, only the sign and direction of the translation t can be estimated and a 3D reconstruction of objects from an image sequence analysis is possible up to a positive scale factor. Since any rotation R in 3D can be described by three parameters and the direction of t by three parameters plus a normalization constraint, e.g. that the Euclidean length $\|t\|$ should be equal to 1, the problem to be solved can be stated as follows:

Given a set $\{(p_{1,i}, p_{2,i}) \mid i = 1 \ldots N\}$ of corresponding point pairs in two perspective images,

estimate the six parameters (subject to a normalization constraint) describing the ego-motion of the camera in terms of a rotation R and a translation t.

3.2 Interpretation of Image Point Correspondences in the Calibrated Case

In the calibrated case, the viewing ray corresponding to any image point can be computed from the inverse projective transformation (see section 2.3). In the following, the viewing rays r_1 and r_2 of a pair of corresponding image points taken from two subsequent frames, the two positions of the optical center of the camera being located at c_1 and c_2, are considered:

$$\begin{aligned} r_1 &= \{m \in I\!R^3 \mid m = c_1 + \lambda d_1, \lambda \in I\!R, \lambda \geq 0\}, \\ r_2 &= \{m \in I\!R^3 \mid m = c_2 + \sigma d_2, \sigma \in I\!R, \sigma \geq 0\}. \end{aligned}$$

Without loss of generality, the direction vectors d_1 and d_2 as well as the displacement $c_2 - c_1$ are assumed to be normalized.

If the two viewing rays r_1 and r_2 had a point of intersection, the error contribution of the pair would be zero. Since this is not the case in general, a 3D object point m has to be constructed using a suitable error model. According to section 2.3, perspectively invariant error cones $C(r_1, \delta(r_1, m))$ and $C(r_2, \delta(r_2, m))$ are associated with m and the corresponding pair of viewing rays. Since both images are acquired by a single camera, the errors $\delta(r_1, m)$ and $\delta(r_2, m)$ are assumed to be identical, i.e.

$$\delta(r_1, m) = \delta(r_2, m) = \psi.$$

Furthermore, m should be an object point with minimal error $\varepsilon(r_1, r_2)$ for the viewing rays r_1 and r_2. Obviously, this requirement is fulfilled if and only if the aperture angle ψ_0 of the corresponding error cones is the infimum of all aperture angles ψ with the property that the intersection of the cones $C(r_1, \psi)$ and $C(r_2, \psi)$ is not empty. Hence, the error $\varepsilon(r_1, r_2)$ of the viewing rays r_1 and r_2 is defined by

$$\varepsilon(r_1, r_2) = \inf\{\psi \in [0, \pi] \mid C(r_1, \psi) \cap C(r_2, \psi) \neq \emptyset\}.$$

The following theorem describes this desired infimum and simplifies the computation of $\varepsilon(r_1, r_2)$.

Theorem 1 *For two viewing rays*

$$\begin{aligned} r_1 &= \{m \in \mathbb{R}^3 \mid m = c_1 + \lambda d_1, \lambda \in \mathbb{R}, \lambda \geq 0\}, \\ r_2 &= \{m \in \mathbb{R}^3 \mid m = c_2 + \sigma d_2, \sigma \in \mathbb{R}, \sigma \geq 0\}, \end{aligned}$$

with $\|c_2 - c_1\| = 1$ *and* $d_1, d_2 \in S_1(0)$*, let*

$$\begin{aligned} \tilde{d}_1 &= d_1 - \langle d_1 | c_2 - c_1 \rangle (c_2 - c_1), \\ \tilde{d}_2 &= d_2 - \langle d_2 | c_2 - c_1 \rangle (c_2 - c_1), \end{aligned}$$

denote the parallel projection of the vectors d_1 *and* d_2 *onto the plane orthogonal to the displacement of the centers of the cones and define*

$$\begin{aligned} s &= \langle d_1 - d_2 | c_2 - c_1 \rangle, \\ k_1 &= \langle \tilde{d}_1 | \tilde{d}_1 + \tilde{d}_2 \rangle, \\ k_2 &= \langle \tilde{d}_2 | \tilde{d}_1 + \tilde{d}_2 \rangle. \end{aligned}$$

Then,

$$\varepsilon(r_1, r_2) = \inf\{\psi \in [0, \pi] \mid C(r_1, \psi) \cap C(r_2, \psi) \neq \emptyset\}$$

is given by

a) $\frac{1}{2} \sphericalangle (d_1, d_2)$ *for* $s \leq 0$.

b) $\sphericalangle (d_1, c_2 - c_1)$ *for* $s > 0,\ k_1 \leq 0$.

c) $\sphericalangle (d_2, c_1 - c_2)$ *for* $s > 0,\ k_2 \leq 0$.

d) $\sphericalangle (d_1, E_e) = \sphericalangle (d_2, E_e)$ *for* $s > 0,\ k_1 > 0$ *and* $k_2 > 0$.
E_e *denotes the plane through* c_1 *(and* c_2*) with normal vector* $n = (d_1 + d_2) \times (c_2 - c_1)$.

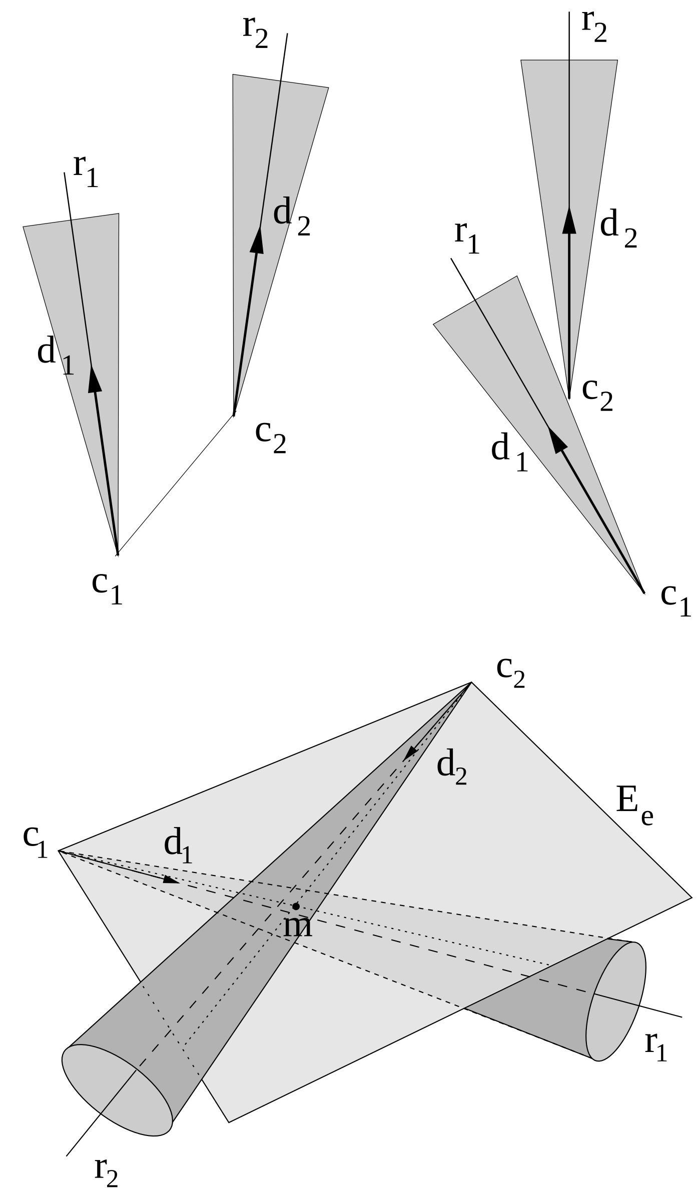

Figure 3: Sample constellations of viewing rays and the corresponding error cones $C(r_1, \varepsilon(r_1, r_2))$, $C(r_2, \varepsilon(r_1, r_2))$ considered in Theorem 1, top left: case a), top right: case b), bottom: case d).

The complete proof of Theorem 1 is given in the appendix of this paper.

Fig. 3 illustrates some typical constellations of viewing rays and the related "minimal" error cones $C(r_1, \varepsilon(r_1, r_2))$ and $C(r_2, \varepsilon(r_1, r_2))$ considered in Theorem 1. Obviously, the intersection of error cones $C(r_1, \psi)$ and $C(r_2, \psi)$ is empty for aperture angles ψ smaller than $\varepsilon(r_1, r_2)$ in the depicted examples and non-empty for larger values of ψ.

The top left image in Fig. 3 represents case a). In contrast to the geometric relations described in part b), c) and d) of Theorem 1, even the intersection of the cones $C(r_1, \varepsilon(r_1, r_2))$ and $C(r_2, \varepsilon(r_1, r_2))$ is empty (except for a displacement of c_1 and c_2 parallel to $d_1 + d_2$) and, therefore, the error $\varepsilon(r_1, r_2)$ is defined by an actual infimum. Hence, the corresponding object point m has to be a point with infinite distance from c_1 and c_2; m can be written as

$$m = \frac{c_1 + c_2}{2} + \lambda(d_1 + d_2), \lambda = +\infty.$$

In case b), which is depicted top right in Fig. 3, the minimum aperture angle $\varepsilon(r_1, r_2)$ is the angle between the direction vector d_1 of the viewing ray r_1 and the displacement $c_2 - c_1$ of the centers of the cones. For this reason, the center c_2 of the cone $C(r_2, \varepsilon(r_1, r_2))$ is contained in the cone $C(r_1, \varepsilon(r_1, r_2))$ and c_2 yields the corresponding 3D object point m with minimal error $\delta(r_1, m) = \delta(r_2, m)$.

An appropriate geometric interpretation for case c) of Theorem 1 is readily achieved by exchanging the roles of the viewing rays r_1 and r_2 in the top right image.

The geometric relations for Theorem 1.d) are illustrated in the bottom example of Fig. 3. The 3D object point m is now a point with finite and non-zero distance from the projection centers c_1 and c_2. This reflects the "regular" situation for viewing rays derived from image point correspondences. The plane E_e is spanned by the vectors c_1, c_2 and m and is a common tangential plane of the cones $C(r_1, \varepsilon(r_1, r_2))$ and $C(r_2, \varepsilon(r_1, r_2))$. Furthermore, the respective error cones lie on different sides of E_e, since

$$\langle (d_1 + d_2) \times (c_2 - c_1) | d_2 \rangle = -\langle (d_1 + d_2) \times (c_2 - c_1) | d_1 \rangle,$$

and, therefore,

$$\langle n | d_2 \rangle = -\langle n | d_1 \rangle.$$

Accordingly, the plane E_e may be regarded as the substitute for the so-called *epipolar plane*, which is mostly used to derive the relevant error contributions for image point correspondences based on a full-perspective modeling.

In the following sections, the results of Theorem 1 are applied to estimate the ego-motion of a single calibrated camera.

3.3 Solution of the Motion Estimation Problem

Suppose that the projection center of the camera, at time τ_1 and τ_2, is located at position c_1 and c_2, respectively. As before, r_1 and r_2 (with normalized direction vectors d_1 and d_2) denote the viewing rays associated with a pair of corresponding image points and m is the related 3D object point (Fig. 4).

In contrast to the previous section, the coordinates of c_1, r_1 and d_1 are now given with respect to the camera coordinate system $\Gamma(\tau_1)$ at time τ_1, while the coordinates of c_2, r_2 and d_2 refer to the camera coordinate system $\Gamma(\tau_2)$ at time τ_2.

Without loss of generality, $\Gamma(\tau_2)$ is chosen as reference system. If $m(\tau_1)$ and $m(\tau_2)$ denote the coordinates of a point m in $\Gamma(\tau_1)$ and $\Gamma(\tau_2)$, respectively, the Euclidean transformation of coordinates is written as

$$m(\tau_2) = R\ m(\tau_1) + t.$$

Accordingly, the coordinate transformation for direction vectors $d(\tau_1)$ and $d(\tau_2)$ is represented by

$$d(\tau_2) = R\ d(\tau_1), \tag{1}$$

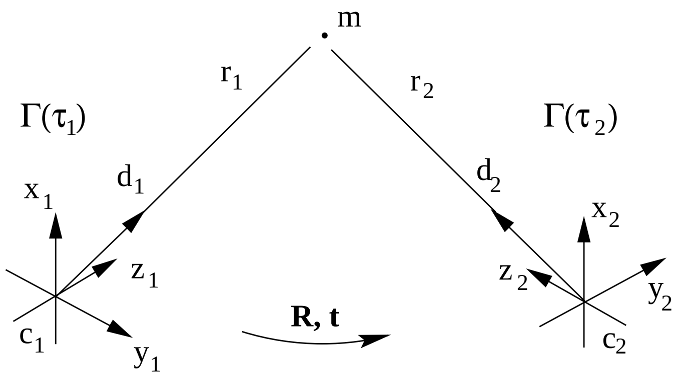

Figure 4: Camera motion model

because of the invariance of directions with respect to translation. Since c_1 and c_2 are the origins of $\Gamma(\tau_1)$ and $\Gamma(\tau_2)$, respectively, their displacement is

$$\begin{aligned} c_2(\tau_2) - c_1(\tau_2) &= c_2(\tau_2) - R\ c_1(\tau_1) - t \\ &= [0,0,0]^T - R\ [0,0,0]^T - t \\ &= -t. \end{aligned} \tag{2}$$

The above mentioned preparations and Theorem 1 now serve as the basis for the formulation of the optimization problem to estimate the ego-motion of the camera. For a set of pairs of corresponding image points

$$\{(p_{1,i},\ p_{2,i}) \mid i = 1 \dots N\},$$

let $d_{1,i} = d_{1,i}(\tau_1)$ and $d_{2,i} = d_{2,i}(\tau_2)$ denote the normalized direction vectors of the viewing rays

$$\begin{aligned} r_{1,i} &= r_{1,i}(\tau_1) &= \{\ \lambda d_{1,i} \mid \lambda \in I\!R, \lambda \geq 0\}, \\ r_{2,i} &= r_{2,i}(\tau_2) &= \{\ \sigma d_{2,i} \mid \sigma \in I\!R, \sigma \geq 0\}, \end{aligned}$$

associated with the image points $p_{1,i}$ and $p_{2,i}$ with respect to $\Gamma(\tau_1)$ and $\Gamma(\tau_2)$, respectively. Substitution of (1) and (2) in the definitions for s, k_1, k_2 and the normal vector n of the plane E_e in Theorem 1 readily yields

$$\begin{aligned} \tilde{d}_{1,i} &= Rd_{1,i} - \langle Rd_{1,i}|t\rangle t, \\ \tilde{d}_{2,i} &= d_{2,i} - \langle d_{2,i}|t\rangle t, \\ s_i &= -\langle Rd_{1,i} - d_{2,i}|t\rangle, \\ k_{1,i} &= \langle \tilde{d}_{1,i}|\tilde{d}_{1,i} + \tilde{d}_{2,i}\rangle, \\ k_{2,i} &= \langle \tilde{d}_{2,i}|\tilde{d}_{1,i} + \tilde{d}_{2,i}\rangle, \\ n_i &= (Rd_{1,i} + d_{2,i}) \times t, \end{aligned}$$

for $1 \leq i \leq N$. Accordingly, the error contribution

$$\varepsilon(p_{1,i}, p_{2,i}) = \varepsilon(r_{1,i}, r_{2,i})$$

for a single point correspondence is

$$\begin{array}{rcll} \frac{1}{2} \sphericalangle (Rd_{1,i}, d_{2,i}), & & \text{for} & s_i \leq 0; \\ \sphericalangle (Rd_{1,i}, -t), & & \text{for} & s_i > 0, k_{1,i} \leq 0; \\ \sphericalangle (d_{2,i}, t), & & \text{for} & s_i > 0, k_{2,i} \leq 0; \\ \sphericalangle (Rd_{1,i}, E_i) & = \sphericalangle (d_{2,i}, E_i), & \text{for} & s_i, k_{1,i}, k_{2,i} > 0. \end{array}$$

Finally, the application of a least squares approach yields the objective function F_0, which has to be minimized to estimate the ego-motion of the camera:

$$F_0 = \sum_{i=1}^{N} \varepsilon(p_{1,i}, p_{2,i})^2.$$

3.4 Qualitative Comparison with Error Functions Derived from Full Perspective Models

For the simple camera model introduced in the example of section 2.2, let P denote the matrix

$$P = A_{\Pi_3} A_{\Pi_2} = \begin{bmatrix} \frac{f}{s_u} & 0 & u_0 \\ 0 & \frac{f}{s_v} & v_0 \\ 0 & 0 & 1 \end{bmatrix}.$$

With this definition,

$$\tilde{p} = P^{-1}[u, v, 1]^T$$

represents the homogeneous coordinates (in the camera coordinate system) of an image point p with pixel coordinates u and v.

The objective function applied in the original papers by Longuet-Higgins [3] and Tsai and Huang [2] is given by

$$F_E = \sum_{i=1}^{N} \varepsilon_E(p_{1,i}, p_{2,i})^2.$$

with

$$\varepsilon_E(p_{1,i}, p_{2,i}) = \langle R\tilde{p}_{1,i} \times t | \tilde{p}_{2,i} \rangle = \langle E\tilde{p}_{1,i} | \tilde{p}_{2,i} \rangle.$$

The 3×3 matrix E in the preceding equation represents the so-called *Essential Matrix*,

$$E = [\ t\ ]_\times R.$$

The *Fundamental Matrix*, introduced by Faugeras, Luong and Maybank in [10], is defined by

$$F = P^{-1^T} E P^{-1} = P^{-1^T} [\ t\ ]_\times R P^{-1}.$$

In these definitions, the mapping $[\ .\ ]_\times : I\!R^3 \to I\!R^{3\times3}$,

$$[\ t\ ]_\times = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix},$$

is used to denote the crossproduct of the vector t and some vector $q \in I\!R^3$ in matrix form, i.e. $t \times q = [\ t\ ]_\times q$.

The improved objective function

$$F_W = \sum_{i=1}^{N} \varepsilon_W(p_{1,i}, p_{2,i}),$$

proposed by Weng, Huang and Ahuja [5, 6], is based on a first order approximation of the optimal weights for the minimization of F_E. Considering a Gaussian distribution of the errors for the extracted pixel coordinates of $p_{1,i}$ and $p_{2,i}$ in the image plane, the following weighted error contribution $\varepsilon_W(p_1, p_2)$ for a single correspondence of image points p_1 and p_2 is derived:

$$\varepsilon_W(p_1, p_2) = \frac{\langle R\tilde{p}_1 \times t | \tilde{p}_2 \rangle^2}{\| R^T(\tilde{p}_2 \times t)\|_{z=0}^2 + \| R\tilde{p}_1 \times t\|_{z=0}^2}.$$

In this context,

$$\|\ [m_x, m_y, m_z]^T\ \|_{z=0} = \sqrt{m_x^2 + m_y^2},$$

denotes the Euclidean norm of the first two coordinates of a vector $m \in I\!R^3$, $m = [m_x, m_y, m_z]^T$.

The error model introduced by Luong and Faugeras [11] uses the Euclidean distances of the image points p_1 and p_2 from the respective epipolar lines for p_2 and p_1 in the image plane. The derived error contribution for a single correspondence (p_1, p_2) is

$$\varepsilon_F(p_1, p_2) = \frac{\langle F p_1 | p_2 \rangle^2}{\| F p_1 \|_{z=0}^2} + \frac{\langle p_1 | F^T p_2 \rangle^2}{\| F^T p_2 \|_{z=0}^2},$$

and the applied objective function is represented by

$$F_F = \sum_{i=1}^{N} \varepsilon_F(p_{1,i}, p_{2,i}).$$

For a qualitative comparison of the individual error terms $\varepsilon(p_1, p_2)^2$, $\varepsilon_W(p_1, p_2)$ and $\varepsilon_F(p_1, p_2)$ associated with the objective functions F_0, F_W and F_F, respectively, the intrinsic parameters of the example pinhole camera are set to $f = s_u = s_v = 1$ and $u_0 = v_0 = 0$. In this case, P is the unit matrix I_3 and $\varepsilon_F(p_1, p_2)$ can be rewritten as

$$\varepsilon_F(p_1, p_2) = \frac{\langle R\tilde{p}_1 \times t | \tilde{p}_2 \rangle^2}{\| R\tilde{p}_1 \times t \|_{z=0}^2} + \frac{\langle \tilde{p}_1 | R^T(\tilde{p}_2 \times t) \rangle^2}{\| R^T(\tilde{p}_2 \times t) \|_{z=0}^2}.$$

The error contributions $\varepsilon(p_1, p_2)$, derived from the proposed conic error model, depend particularly on the respective signs of s, k_1 and k_2 in Theorem 1. From a perspective point of view, two basic geometric properties are reflected in these criteria: if the camera is moving forward (backward), the corresponding point p_2 in the second image is bound to

- have a larger (smaller) distance from the epipole than p_1;
- be located on the same section – with respect to the epipole – of the epipolar line as p_1.

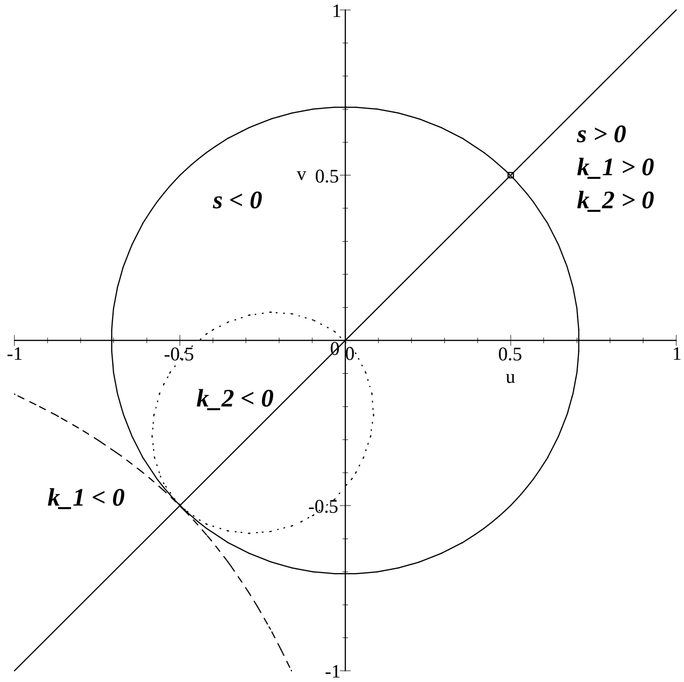

Figure 5: Sample boundaries and respective signs of s (solid), k_1 (dashed) and k_2 (dotted) from Theorem 1 in the image plane for a pair of corresponding image points $p_1 = [0.5, 0.5]^T$ and $p_2 = [u, v]^T$; $R = I_3$, $t = [0, 0, -1]^T$.

A more detailed analysis with respect to the image plane is shown in Fig. 5, which depicts the image point $p_1 = [0.5, 0.5]^T$, its epipolar line and the signs of s, k_1 and k_2 in relation to the position of the image point $p_2 = [u, v]^T$ for $R = I_3$ and $t = [0, 0, -1]^T$. The solid, dashed and dotted curves represent image points p_2 with $s = 0$, $k_1 = 0$ and $k_2 = 0$, respectively.

Fig. 6 and Fig. 7 visualize the relevant error contributions of a pair of corresponding image points $p_1 = [0.5, 0.5]^T$ and $p_2 = [u, v]^T$ with respect to the position of p_2 for a pure translation of the camera along the optical axis, i.e. $R = I_3$ and $t = [0, 0, \pm 1]^T$. In this example, the epipole is located at $[0, 0]^T$ and the epipolar line for p_1 is exactly the main diagonal.

For the proposed conic error model, the error contributions $\varepsilon(p_1, p_2)$ differ significantly for $t = [0, 0, 1]^T$ and $t = [0, 0, -1]^T$ (Fig. 6). In both cases, only points p_2 on the epipolar line with a feasible 3D interpretation in front of the camera yield $\varepsilon(p_1, p_2) = 0$, i.e. image points p_2 with pixel coordinates u and v satisfying $u = v \geq 0.5$ for $t = [0, 0, -1]^T$, and $0 \leq u = v \leq 0.5$ for $t = [0, 0, 1]^T$.

In contrast to this approach, the error contributions $\varepsilon_W(p_1, p_2)$ and $\varepsilon_F(p_1, p_2)$ (Fig. 7) derived from full-projective models remain unchanged if the sign of the translation is altered. All possible correspondences p_2 along the epipolar line for p_1, for instance, are treated identically with error

$$\varepsilon_W(p_1, p_2) = \varepsilon_F(p_1, p_2) = 0.$$

Hence, no distinction is made if the corresponding 3D object point m is located behind the camera.

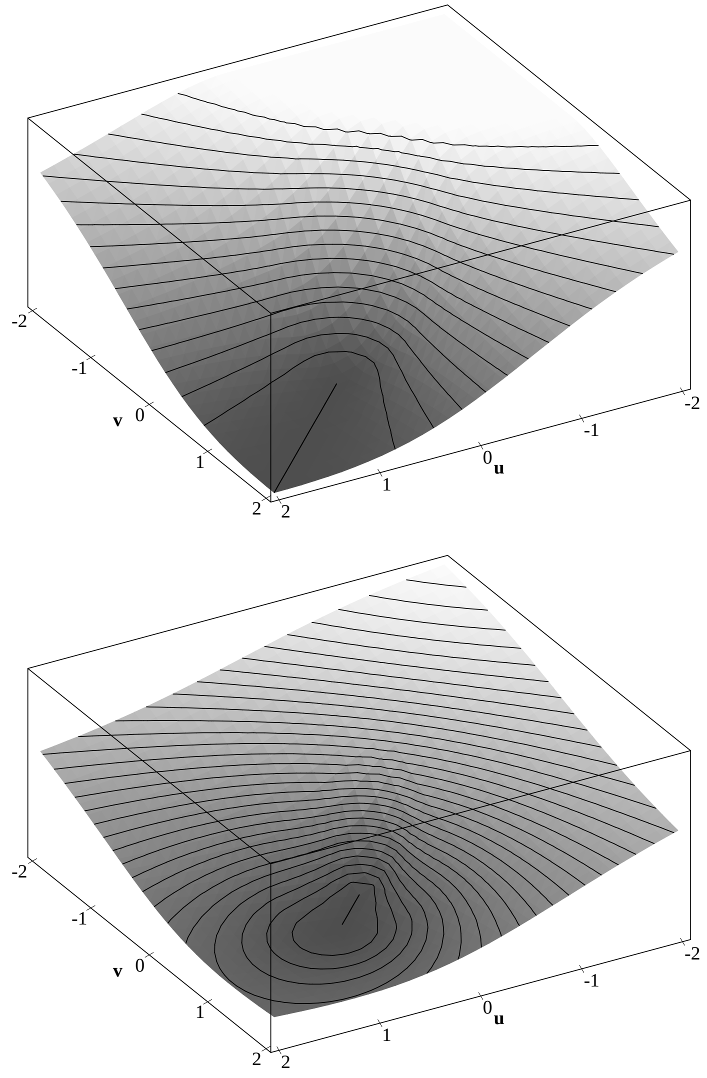

Figure 6: Squared error contribution $\varepsilon(p_1, p_2)^2$ for a pair of corresponding image points $p_1 = [0.5, 0.5]^T$ and $p_2 = [u, v]^T$; $R = I_3$, $t = [0, 0, -1]^T$ (top), $t = [0, 0, 1]^T$ (bottom).

Note that $\varepsilon_F(p_1, p_2)$ is undefined if $R\tilde{p}_1$ or $\tilde{p}_2$ is parallel to the translation t of the camera, which results in numerical instabilities for image points very close to the relevant epipole. This discontinuity can be seen in the plot for $\varepsilon_F(p_1, p_2)$ near $u = v = 0$.

3.5 Possible Simplifications and Initial Value Estimation

In order to use ego-motion analysis with automotive applications, a test platform including

- an image sensor,
- a high performance image processing computer and
- various sensors detecting the vehicle motion (velocity, acceleration, suspension-deflection and gyro sensors)

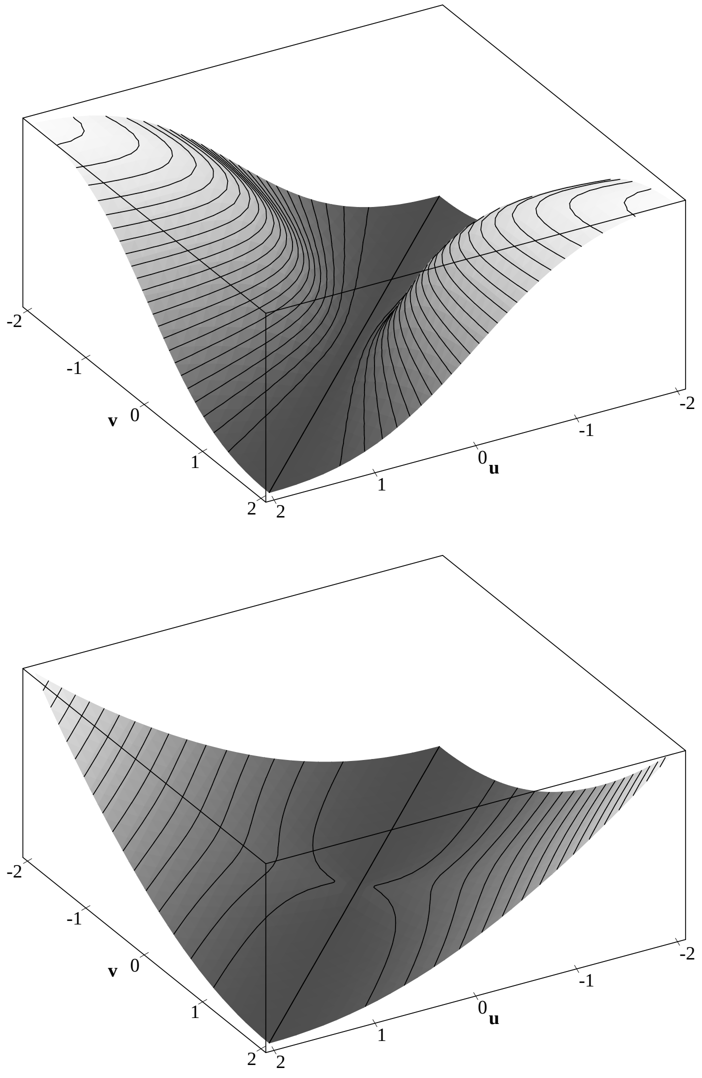

Figure 7: Error contributions $\varepsilon_W(p_1, p_2)$ (top) and $\varepsilon_F(p_1, p_2)$ (bottom) for a pair of corresponding image points $p_1 = [0.5, 0.5]^T$ and $p_2 = [u, v]^T$; $R = I_3$, $t = [0, 0, \pm 1]^T$.

has been installed in a passenger car. The image sensor is a CCD camera, mounted behind the windshield near the rear-view mirror. The digital frame format used as input for the image processing machine features 512×256 pixels per frame at a frame rate of 50 Hz and 8-bit grey-value per pixel.

An angular parameterization of the rotation matrix R is used for the non-linear minimization problem, which is associated with the objective function derived in the last section, since the rotational components related to the three main axes of the test car are of particular interest. The matrices R_x, R_y and R_z for rotations around the x-, y- and z-axis by ϕ_x, ϕ_y and ϕ_z, respectively, are given by

$$R_x(\phi_x) \quad = \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\phi_x) & \sin(\phi_x) \\ 0 & -\sin(\phi_x) & \cos(\phi_x) \end{bmatrix},$$

$$R_y(\phi_y) = \begin{bmatrix} \cos(\phi_y) & 0 & -\sin(\phi_y) \\ 0 & 1 & 0 \\ \sin(\phi_y) & 0 & \cos(\phi_y) \end{bmatrix},$$

$$R_z(\phi_z) = \begin{bmatrix} \cos(\phi_z) & \sin(\phi_z) & 0 \\ -\sin(\phi_z) & \cos(\phi_z) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Any general rotation R in 3D may be described as a product of these three matrices:

$$R = R_z(\phi_z) \cdot R_y(\phi_y) \cdot R_x(\phi_x).$$

Since it is impossible to determine the absolute value of the translation between consecutive frames, a normalization constraint for the three components t_x, t_y and t_z of the translation vector t has to be considered. In general, one can either make use of the restriction $\|t\| = 1$ or apply an angular parameterization such as

$$t = [\sin\alpha \sin\beta, \cos\alpha \sin\beta, \cos\beta]^T.$$

In the first case, an approximation problem with six parameters and one non-linear constraint has to be solved. In the second situation, a non-linear approximation problem results, comprising five parameters without constraints.

In comparison, self-calibration methods based on the computation of the *Fundamental Matrix* require the minimization of a non-linear objective function depending on at least seven independent parameters. This implies a considerable increase of computational expense, which may be prohibitive in real-time applications. Moreover, it is not possible to derive the motion parameters of the camera from a single *Fundamental Matrix*. In order to obtain quantitatively reliable motion estimates in the Euclidean space, several *Fundamental Matrices* have to be estimated; subsequently the intrinsic parameters of the camera have to be extracted and only then the ego-motion of the camera can be determined. This requires additional computational effort and results in considerable (and therefore often undesired) temporal delays.

The experiments described in section 4.1 show that a minimum of 50 to 100 corresponding viewing rays is required to achieve superior accuracy for our application. In this case, the optimization of the objective function F_0 takes between 0.005 and 2.8 seconds on a SUN workstation equipped with a 50 MHz SuperSparc CPU, depending on the quality of the supplied initial values in relation to the requested precision of the parameters to be estimated.

Regarding the inherent real-time requirements of our system (frame repetition rate is 50 Hz and maximum total delay time is 20 msec.), it is not possible to ensure the termination of the optimization of the objective function in due time. Since the restriction of the minimization process to a fixed number of iterations is crucial with respect to the accuracy of the achieved results, the following additional measures may be taken to reduce the computation time.

- Instead of minimizing the aperture angles $\varepsilon(r_1, r_2)$ of the error cones for the viewing rays r_1 and r_2 with direction vectors $d_1, d_2 \in S_1(0)$, $\varepsilon'(r_1, r_2) = \sin(\varepsilon(r_1, r_2))$ is minimized. Since the typical size of the aperture angle corresponds to less than 1 pixel (i.e. $0.1°$), this approximation is uncritical with respect to the accuracy of the estimated motion parameters.

For a normalized translation vector t, $\|t\| = 1$, the individual squared error contributions $\varepsilon_i'(r_{1,i}, r_{2,i})^2$ are then given by

$$\begin{array}{ll} \dfrac{\|Rd_{1,i} - d_{2,i}\|^2}{4}, & \text{for} \quad s_i \leq 0; \\ 1 - \langle Rd_{1,i} | t\rangle^2, & \text{for} \quad s_i > 0, k_{1,i} \leq 0; \\ 1 - \langle d_{2,i} | t\rangle^2, & \text{for} \quad s_i > 0, k_{2,i} \leq 0; \\ \dfrac{\langle Rd_{1,i} \times t | d_{2,i}\rangle^2}{\|(Rd_{1,i} + d_{2,i}) \times t\|^2}, & \text{for} \quad s_i, k_{1,i}, k_{2,i} > 0, \end{array}$$

and the objective function is simplified to

$$F_1 = \sum_{i=1}^{N} \varepsilon_i'(r_{1,i}, r_{2,i})^2.$$

- The viewing rays $r_{1,i}$ and $r_{2,i}$ are replaced by viewing lines

$$\begin{aligned} l_{1,i} &= \{ \lambda d_{1,i} \mid \lambda \in I\!R\}, \\ l_{2,i} &= \{ \sigma d_{2,i} \mid \sigma \in I\!R\}, \end{aligned}$$

  and the error cones $C(r, \psi)$ by "double" cones

$$\mathcal{C}(l, \psi) = C(r, \psi) \cup C(-r, \psi) = C_{c,d,\cos(\psi)} \cup C_{c,-d,\cos(\psi)}.$$

  The resulting objective function, which corresponds to a full-projective modeling of the motion estimation problem with a Gaussian distribution of the errors for the direction vectors $d_{1,i}$ and $d_{2,i}$ in the unit sphere $S_1(0)$, is given by

$$F_2 = \sum_{i=1}^{N} \frac{\langle Rd_{1,i} \times t | d_{2,i}\rangle^2}{\|(Rd_{1,i} + d_{2,i}) \times t\|^2}.$$

- The modeling of errors is restricted to a single image. That is, the error cones $\mathcal{C}(l_1, \psi)$ and $\mathcal{C}(l_2, \psi)$ are replaced by $\mathcal{C}(l_1, \psi')$ and $\mathcal{C}(l_2, 0)$. Accordingly, the objective function becomes

$$F_3 = \sum_{i=1}^{N} \frac{\langle Rd_{1,i} \times t | d_{2,i}\rangle^2}{\|d_{2,i} \times t\|^2}.$$

- In our application, the camera is rigidly attached to the test car. Taking advantage of camera calibration, it is possible to determine the relative location and orientation of the camera coordinate system in relation to that of the car. Therefore, the direction of the longitudinal axis of the car t_0 may be used as a reasonable estimate for the direction of the translational motion vector of the camera. Then, the objective function is defined by

$$F_4 = \sum_{i=1}^{N} \frac{\langle d_{2,i} \times t_0 | Rd_{1,i}\rangle^2}{\|d_{2,i} \times t_0\|^2}.$$

- Due to the high frame repetition rate of 50 Hz, the orientation of the camera changes only very little between two frames. Consequently, the angles ϕ_x, ϕ_y and ϕ_z are very small (i.e. never larger than 1°). The affine approximations

$$\cos(\alpha) \approx 1,$$
$$\sin(\alpha) \approx \alpha,$$

for cos– and sin–terms and neglecting higher order terms yield the following approximation R_0 for the rotation matrix R

$$R_0 = \begin{bmatrix} 1 & \phi_z & -\phi_y \\ -\phi_z & 1 & \phi_x \\ \phi_y & -\phi_x & 1 \end{bmatrix}.$$

The objective function thus becomes

$$F_5 = \sum_{i=1}^{N} \frac{\langle d_{2,i} \times t_0 | R_0 d_{1,i} \rangle^2}{\| d_{2,i} \times t_0 \|^2}.$$

The minimization problems associated with the objective functions F_1, F_2, F_3 and F_4 are still non-linear. However, since the rotation matrix R does no longer occur in the denominators of the error terms in F_3 and F_4, and since the direction of the translation is estimated from the calibration process for F_4, more efficient solution algorithms with higher numerical stability can be applied. The optimization of the objective function F_5 can be carried out by standard linear least-squares techniques. Therefore, F_5 is well-suited for generating initial values of the rotational angles ϕ_x, ϕ_y and ϕ_z, required as input for the iterative minimization algorithms for F_0, F_1, F_2, F_3 or F_4.

4 Experimental Results

The algorithmic approach presented in this paper has been tested by various experimental investigations. This section is devoted to an analysis and discussion of the corresponding results. The numerical stability as well as the accuracy of the motion estimates have been studied with both synthetic and natural image data (see subsections 4.1 and 4.2, respectively).

In all experiments presented for image sequences, the following convention is applied: the estimation results for a frame j refer to the motion of the camera between the acquisition of the images $j-1$ and j.

4.1 Experiments Using Synthetic Data

In order to analyze the dependence of the robustness and accuracy of the achieved estimation on the quality of input data, simulated camera motion trajectories have been applied to the optimization process. Specifically, the following influences have been investigated:

- the precision of the input point correspondences,

- the accuracy of the supplied initial values (convergence towards the global minimum of the objective function),
- the depth variance of the object points and
- the total number of input point correspondences.

The experiment described at the end of this subsection uses a flat-shaded computer-animated image sequence of a road scene; this introduces quantization noise and unknown depth variance. The ego-motion results are compared with those of real data presented in section 4.2.

4.1.1 Computer Simulated Camera Motion

A standard pinhole camera with focal length $f = 8 \cdot 10^{-3}$ and an image resolution of 512×512 pixels is simulated. The pixels are assumed to be arranged on a quadratic grid with a grid constant of $1.15 \cdot 10^{-5}$. Image point correspondences are obtained by

- random generation of a set of N points uniformly distributed in a cube C located in front of the camera,
- displacement of the camera to a new position combining a rotation R with a subsequent translation t,
- transformation of the 3D point coordinates to the camera coordinate system corresponding to the new position,
- projection of the 3D object points to the image planes related to both camera positions and
- addition of uniformly distributed noise with maximum amplitude ϱ to the x- and y-coordinates of the projected points in both images.

A typical sample of simulated image point correspondences is shown in Fig. 8.

For all experiments in this subsection, the initial values used for the iterative minimization of the non-linear objective function F_0 have been set to $\tilde{t} = [0, 0, -1]^T$, $\tilde{\phi}_x = 0$, $\tilde{\phi}_y = 0$ and $\tilde{\phi}_z = 0$.

Since the absolute value of the translation vector cannot be recovered from the motion estimation process, the angular difference

$$\Delta t = \sphericalangle\,(\hat{t}, t)$$

between the estimated translational direction t and the exact translation $\hat{t}$ is taken as error measure Δt. The estimates ϕ_x, ϕ_y and ϕ_z for the rotational angles are compared to the exact values $\hat{\phi}_x$, $\hat{\phi}_y$ and $\hat{\phi}_z$, respectively, by their absolute deviations, e.g.

$$\Delta\phi_x = \mid \hat{\phi}_x - \phi_x \mid.$$

In a first sequence of tests, only the absolute amount of noise applied to the image coordinates of the point correspondences has been varied and the stability of the resulting motion estimates has been analyzed.

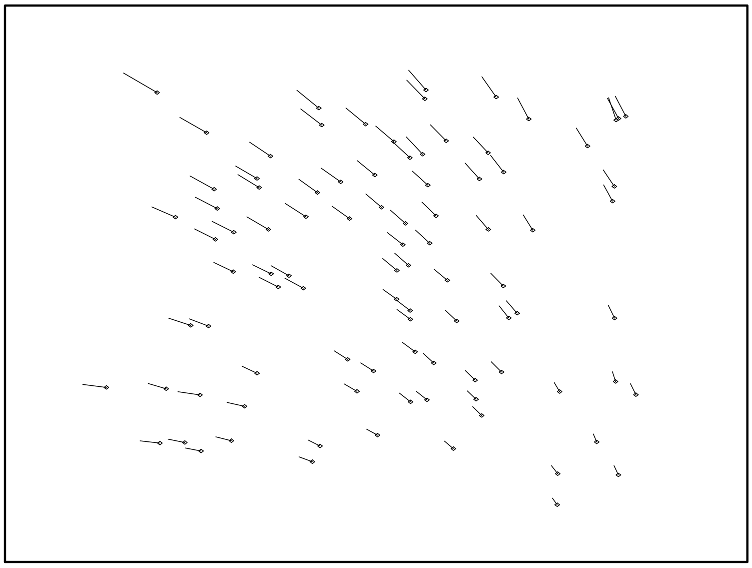

Figure 8: A sample of simulated image point correspondences; the dots depict the projected positions of the generated object points in the first image and the attached line segments visualize the displacements of the respective projections in the second image

With the actual experiments, a pure translation along the optical axis of the camera towards the cube

$$C = [-10; 10] \times [-10; 10] \times [20; 50]$$

has been simulated, i.e. the translation vector has been $\hat{t} = [0, 0, -1]^T$ and the rotational angles have been set to 0. The noise amplitude ϱ has been increased for both image directions from 0 to 4 pixels in steps of 0.1. For each noise amplitude, 100 tests have been carried out, using 100 randomly generated points taken from the cube C. The average absolute errors of the estimated motion parameters are presented in Fig. 9.

The figure shows that the average errors of the estimated translational direction and of the three rotational angles increase approximately linearly with the applied noise amplitude ϱ.

It is emphasized that the image displacement of a pair of corresponding points is invariant to a simultaneous increase of the translational length and the depth of the related 3D object point. Vice versa, in case of a pure translation of the camera, the accuracy of the estimates is improved if either the noise amplitude is reduced or the size of the image displacement is increased. For a given noise amplitude, this means that the utilization of foreground object points or application of a "longer" translation will improve the accuracy of the results by a corresponding factor.

In automotive applications, the absolute value of the translation could only be varied via the frame repetition rate. In practice this is not feasible, because of the huge displacements of all point correspondences in case of large rotational rates and of the projections with foreground object points at high speed. In addition, considering the integration into a real-time computer vision system, ego-motion analysis usually precedes the classification of objects. For this reason, there is no a-priori knowledge about the depth of particular object points. In contrast to many indoor applications, the only way

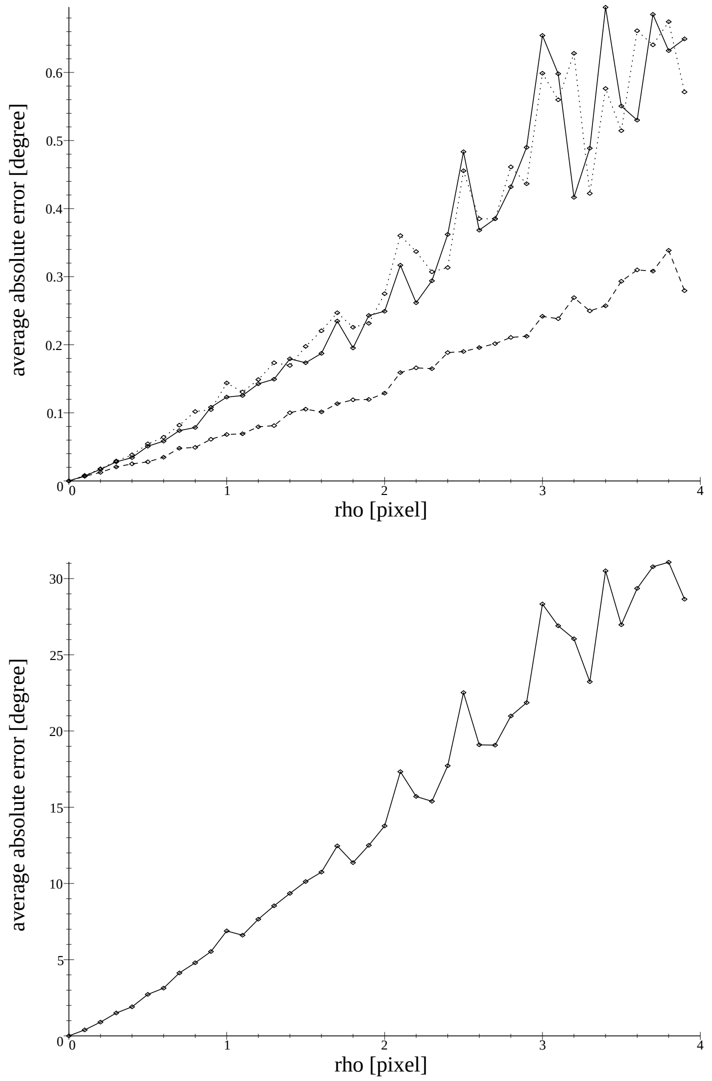

Figure 9: Average absolute errors of the estimated motion parameters in relation to the maximum noise amplitude ϱ, $\Delta\phi_x$ (solid), $\Delta\phi_y$ (dotted), $\Delta\phi_z$ (dashed) (top) and Δt (bottom)

to achieve sufficiently accurate results is therefore to compute the point correspondences with sub-pixel precision and to apply robust and precise error minimization procedures.

In the second series of experiments, the camera motion has been generalized to include translational components perpendicular to the camera axis as well as rotational transitions in order to study the dependence of the motion estimation errors on the details of the camera trajectory. Furthermore, it has been investigated whether the set of used initial values is sufficient to ensure convergence towards the global minimum of the objective function F_0.

The maximum noise amplitude has been set to 1 pixel. The parameterization

$$t = [\sin\alpha\sin\beta, \cos\alpha\sin\beta, -\cos\beta]^T$$

has been used for the translation of the camera. The angles ϕ_x, ϕ_y, ϕ_z, α and β have been chosen randomly from an interval $[-\Theta, \Theta]$. 28 different values for Θ have been applied ranging from 0° to 12.15°. For every distinct value of Θ, 100 experiments have been performed. As before, 100 object points have been randomly selected from the volume of the cube

$$C = [-10; 10] \times [-10; 10] \times [20; 50]$$

for each experiment. The average absolute errors of the estimated motion parameters are shown in Fig. 10.

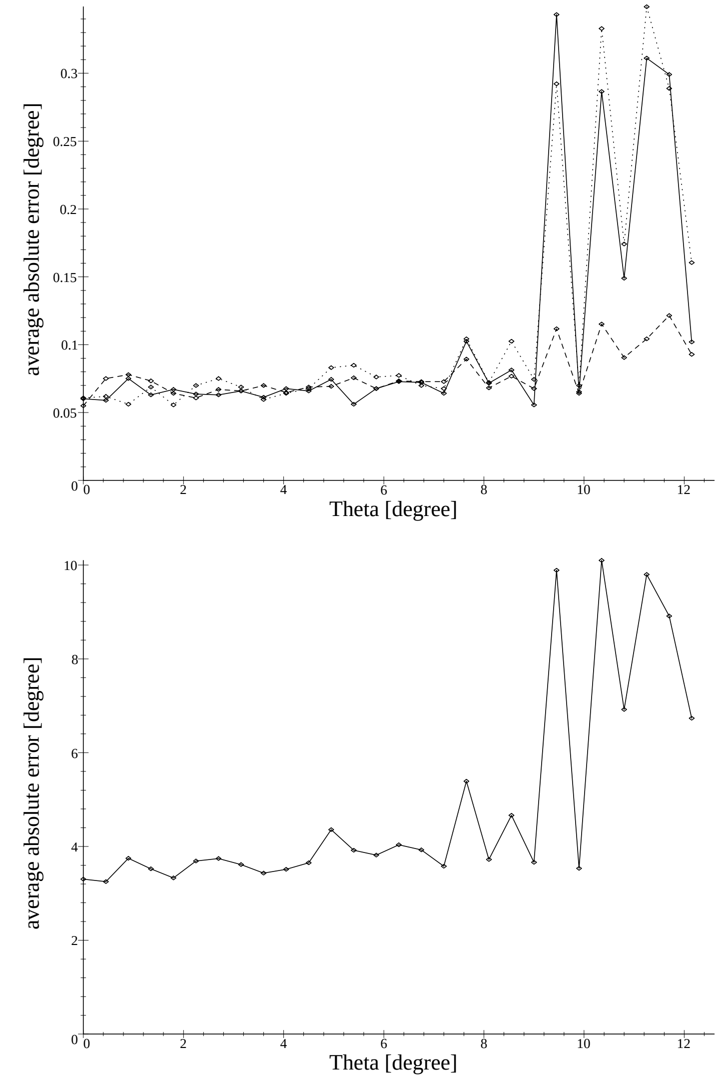

Figure 10: Average absolute errors of the estimated motion parameters in relation to the size of Θ using default initial values. $\Delta\phi_x$ (solid), $\Delta\phi_y$ (dotted), $\Delta\phi_z$ (dashed) (top) and Δt (bottom)

It can be seen that the absolute errors of the estimated parameters related to the rotation and the direction of translation are approximately independent of Θ, as long as Θ is smaller than 7°. In our study, we have observed that this holds for even larger values of Θ if the initial values are adapted to the actual camera motion. Therefore, in the experiments discussed here, the optimization procedure tends to converge towards a local minimum of the non-linear objective function F_0 for values of Θ larger than 7°.

On the other hand, these experiments show that starting values for the angles ϕ_x, ϕ_y, ϕ_z and t with an accuracy as poor as $7°$ still yield the global minimum and the correct motion estimation. It is therefore not a problem to ensure proper convergence behavior since starting values with the requested accuracy can be easily generated by standard heuristics in all computer vision applications the authors of this paper are aware of.

The third series of experiments investigates the stability of the motion estimation for a given number of point correspondences with respect to the variation of the depth of 3D object points in relation to the length of the translation vector.

This experiment is motivated by algorithmic problems typical for vehicle applications mostly dealing with 3D object points with widely varying distances. In this case, a perspectively invariant error modeling is mandatory to ensure the proper weighting of object points according to their distances from the camera. With fast translational motion the matching of image points is considerably complicated by big displacements of the projections of foreground object points. On the other hand, for image points close to the focus of expansion the correspondence search is generally successful. Thus, matched point pairs tend to occur near the relevant epipole, in which case the determination of the ego-motion is (partially) unstable.

The maximum noise amplitude ϱ for these experiments has been limited to 1 pixel and the motion parameters of the camera have been set to $t = [0.02, 0.03, -1.0]^T$, $\phi_x = 1°$, $\phi_y = -1°$ and $\phi_z = 1°$. The variance of the relative depth of 3D object points has been modeled by a random selection of points from cubes of variable depth:

$$C = [-10; 10] \times [-10; 10] \times [20; z_{\max}].$$

The maximum distance $z_{\max}$ has been increased in steps of 10 starting at $z_{\max} = 40$ up to $z_{\max} = 430$. For every distinct value of $z_{\max}$, 40 experiments have been carried out, using 100 randomly generated object points for each experiment. The average absolute errors of the estimated motion parameters are depicted in Fig. 11.

The figure shows that the average errors of the results for ϕ_x and ϕ_y depend clearly less on the relative depth variance of the 3D object points than that of ϕ_z and t. For values of $z_{\max}$ between 100 and 250 the accuracy of the estimates for ϕ_x and ϕ_y remains approximately unchanged.

In contrast, the precision of the results for the angle ϕ_z as well as for the translational direction is affected significantly by the fact that most of the extracted point correspondences are located close to the epipole.

This is due to the fact that the direction of the optical axis associated with ϕ_z is almost aligned with the translational motion of the camera. As a consequence, errors of the estimates for t also lead to errors for the corresponding angle ϕ_z.

Note that the average errors for ϕ_x, ϕ_y and t increase if the value of $z_{\max}$ becomes very small. This is caused by fundamental instabilities in the estimation of general camera motions for degenerated sets of object points taken from a plane (particularly for planes perpendicular to the translation vector t).

In the last experiment using synthetic data, the precision of the motion estimation results has been examined in relation to the number of point correspondences. The simulation setup has been almost identical to that of the first experiment with the exception that the maximum noise amplitude has been set to 1.0 pixel and the number of point correspondences has been varied between 20 and 210.

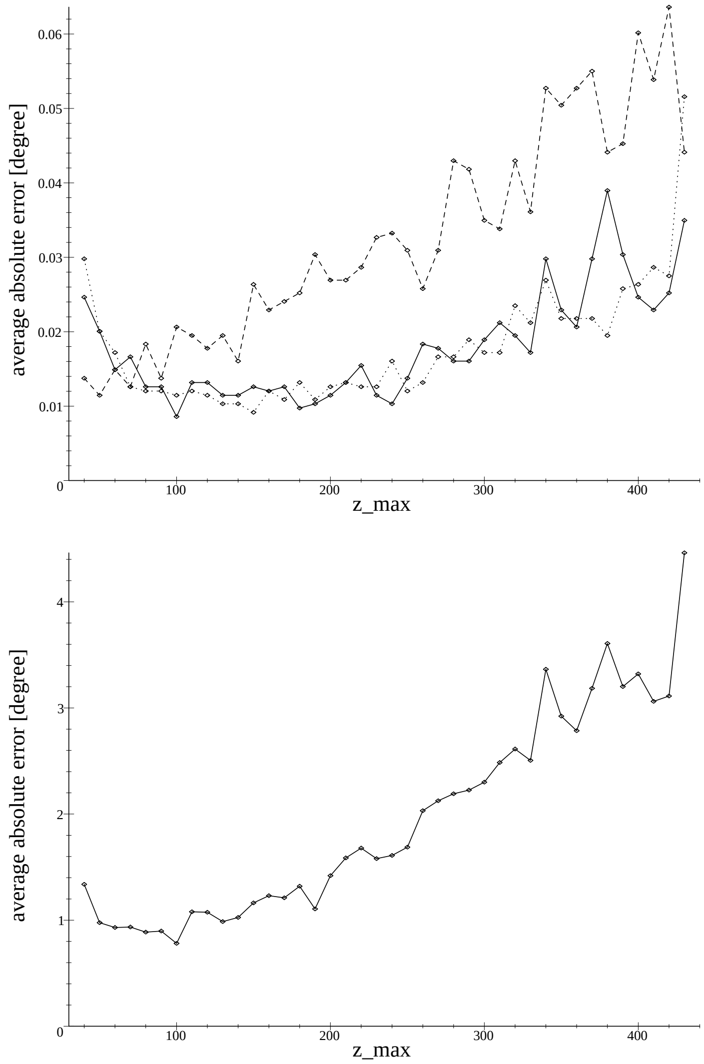

Figure 11: Average absolute errors of the estimated motion parameters in relation to the variation of relative depth of the generated object points, $\Delta\phi_x$ (solid), $\Delta\phi_y$ (dotted), $\Delta\phi_z$ (dashed) (top) and Δt (bottom)

The average absolute errors of the corresponding motion estimation results are shown in Fig. 12.

It can be seen that a good precision of the estimated motion parameters (i.e. average errors of less than $0.02°$ for the rotation angles ϕ_x, ϕ_y, ϕ_z and of less than $1°$ for the translational direction) requires some 50 to 150 point correspondences. The use of more point correspondences yields no significant improvement of the estimates. In practical applications, there is certainly a tradeoff to be made, since the number of high-accuracy correspondences is limited and the computational expense of the optimization procedure increases. For natural image sequences, a total number of 50 to 100 point correspondences has turned out to yield stable and sufficiently accurate results for our application.

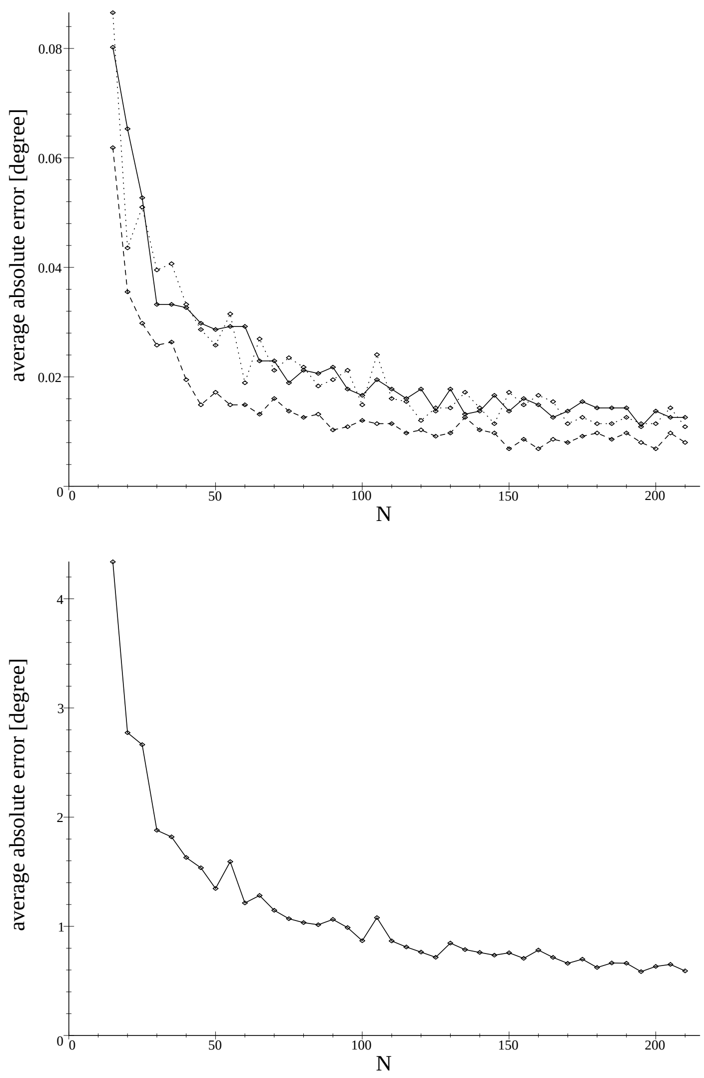

Figure 12: Average absolute errors of the estimated motion parameters in relation to the number N of point correspondences, $\Delta\phi_x$ (solid), $\Delta\phi_y$ (dotted), $\Delta\phi_z$ (dashed) (top) and Δt (bottom)

4.1.2 Experiments Using Computer Animated Road Scenes

The experiment of this subsection combines point correspondences extracted from an animated, flat-shaded scene with simulated motion trajectories. A sample image of a computer-animated road sequence is shown in Fig. 13.

With this experiment,

- the intrinsic parameters of the camera,
- the exact 3D positions of all objects in the scene and
- the camera motion parameters

are known exactly.

Figure 13: A sample image of a computer-animated road sequence

Since the point correspondences are now extracted by image processing algorithms, they are subject to quantization noise and potential mismatch.

The robustness of the motion estimation has been examined repeatedly for all images of various sequences. Results obtained from one of the animated road scenes are presented in Table 1. The differences between the estimated rotational angles and their exact values used by the graphics animation procedure are listed for 30 subsequent frames. The maximum absolute error is typically smaller than $0.05°$, the average absolute error lies between $0.01°$ and $0.02°$.

4.2 Experiments with Natural Data

The preceding subsection was devoted to the verification and detailed stability analysis of our approach. In the following, it is applied to the more challenging task of motion estimation from natural image sequences recorded with the test car, which was equipped with gyro sensors capable of monitoring the rotational motion of the car. Fig. 14 shows two image samples of a natural road scene.

Subsequently, the following experiments will be discussed: first, the precision of the estimated rotational motion parameters derived from the conic error model (i.e. achieved by the minimization of the objective function F_0) is compared to the results of the approach described by Weng, Huang and Ahuja (F_W) and the angular rates as measured by the gyro-sensors; second, the error functions F_0 and F_W are tested with respect to the proper detection of outliers and the correct estimation of the ego-motion of the camera for an image sequence with different relative motions; third, the stability of the "calibrated approaches" F_0 and F_W is investigated in comparison with results derived from the computation of the *Fundamental Matrix* and the objective function F_F in the uncalibrated case; finally, the influence of all approximations described in section 3.5 is investigated by a comparison of the results obtained from the simplified model (F_5) with those of the full-scale approaches F_0 and F_W.

4.2.1 Comparison with Data from Gyro-Sensors

In Fig. 15, the three rotational angles measured by the gyro-sensors are compared to the results of our motion estimation approach (F_0) and the respective results achieved with the objective function F_W proposed by Weng, Huang and Ahuja [5, 6]. The dotted lines

Table 1: Differences between the exact values of $\hat{\phi}_x$, $\hat{\phi}_y$, $\hat{\phi}_z$ and the corresponding estimates ϕ_x, ϕ_y, ϕ_z for the computer-animated road scene

Frame	$\hat{\phi}_x - \phi_x$	$\hat{\phi}_y - \phi_y$	$\hat{\phi}_z - \phi_z$
1	-0.001°	-0.014°	0.018°
2	0.001°	0.014°	0.013°
3	0.010°	0.015°	0.001°
4	-0.003°	-0.011°	0.011°
5	-0.001°	0.014°	0.015°
6	-0.001°	0.017°	0.019°
7	-0.011°	0.002°	0.007°
8	0.009°	-0.001°	0.001°
9	-0.001°	0.011°	0.007°
10	-0.002°	0.013°	-0.006°
11	-0.003°	-0.008°	-0.003°
12	0.002°	0.015°	0.025°
13	-0.006°	0.014°	0.014°
14	-0.003°	-0.013°	0.020°
15	-0.002°	0.016°	0.022°
16	-0.003°	0.014°	-0.005°
17	0.004°	0.003°	0.005°
18	-0.003°	0.009°	0.021°
19	-0.001°	0.018°	0.036°
20	-0.006°	0.017°	0.023°
21	-0.026°	-0.014°	0.005°
22	-0.030°	0.017°	0.023°
23	0.013°	0.011°	-0.005°
24	0.018°	-0.016°	-0.010°
25	0.045°	0.009°	0.000°
26	0.009°	0.012°	0.005°
27	-0.005°	-0.002°	0.013°
28	-0.018°	-0.002°	0.005°
29	-0.004°	0.008°	-0.016°
30	-0.007°	0.004°	-0.015°
max. abs. error	0.045°	0.018°	0.036°
av. abs. error	0.008°	0.011°	0.012°

represent the gyro sensor data, while the solid lines and the dots depict the corresponding rotational angles obtained by the minimization of F_0 and F_W, respectively. The figure shows that the differences between the results of F_0 and F_W are extremely small. Furthermore, the maximum absolute deviation of the estimates from the respective data from gyro sensors is less than 0.05°. Accordingly, all data are correlated very well. Moreover, we emphasize that the gyro sensor data has been subject to intrinsic temporal filtering, in contrast to the results of motion estimation from image analysis, which have been not.

Figure 14: Two sample frames of an image sequence of a natural road scene, Frame 1 (top) and Frame 100 (bottom)

4.2.2 Iterative Refinement of the Estimated Ego-Motion

An iterative improvement of the estimation results is necessary in case of

- potential mismatches (e.g. in similarly textured image regions) or
- image correspondences that belong to different relative motions (e.g. other cars in vehicle applications).

A standard technique to distinguish the distinct correspondences in order to refine the accuracy of the estimated parameters is the so-called detection of outliers. In this case, point correspondences with large individual error contributions to the applied objective function are temporarily removed from the data set and the estimation process is repeated for the remaining correspondences.

Fig. 16 shows a sample frame taken from an image sequence of a road scene with two cars in front the camera. Accordingly, the extracted image point correspondences belong to three different sets of motion parameters: the movement of the camera with respect to the ground and the two relative movements in relation to the cars. For the accurate estimation of the ego-motion (i.e. with respect to the ground), it is necessary to distinguish the different motions and to group the pairs of corresponding image points into distinct subsets.

In the present experiment, the motion parameters have been first estimated for all extracted point correspondences. Then, point pairs with an individual error contribution of more than 60% of the maximum error value among all correspondences have been removed from the estimation process and a new set of motion parameters has been

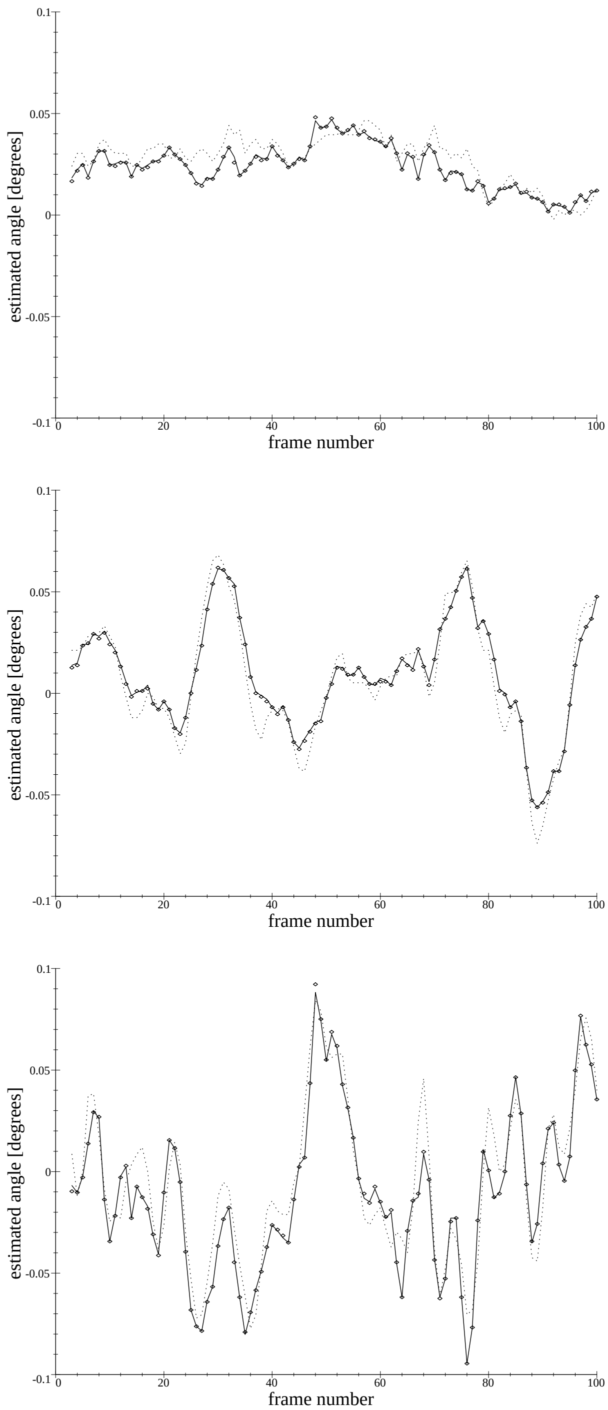

Figure 15: Estimation results for the angles ϕ_x (top), ϕ_y (mid) and ϕ_z (bottom); based on the minimization of the objective functions F_0 (solid lines) and F_W (dots), the respective data from gyro-sensors is depicted in dotted lines

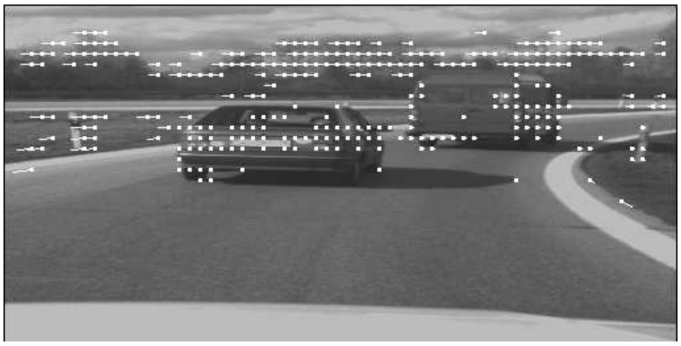

Figure 16: A sample frame of an image sequence of a natural road scene with moving objects and the extracted image point correspondences, the white squares depict the image positions of the image points in the previous frame and the attached line segments visualize their displacements with respect to the actual image

calculated based on the remaining correspondences. This procedure has been iterated five times.

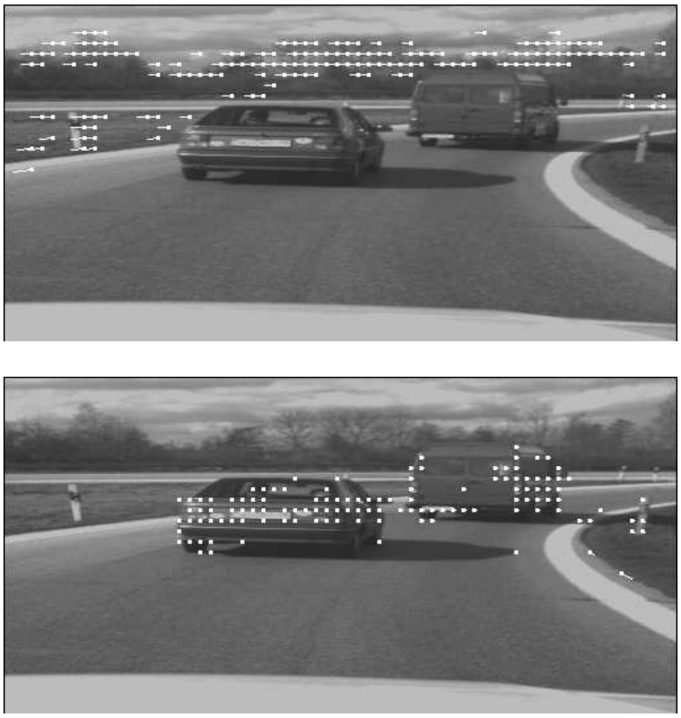

Figure 17: Selected image point correspondences (top) and detected outliers (bottom) after an iterative refinement of the motion estimation by the proposed objective function F_0.

Fig. 17 and Fig. 18 show the derived two subsets of image point correspondences based on the minimization of the proposed objective function F_0 and the objective function F_W, respectively. For F_0 a clear separation of point correspondences related to moving and non-moving objects in the scene is achieved. In contrast, the objective function F_W fails

Figure 18: Selected image point correspondences (top) and detected outliers (bottom) after an iterative refinement of the motion estimation by the objective function F_W.

to correctly distinguish the different motion parameters and a "spoiled" estimation of the ego-motion of the camera results.

Fig. 19 shows the corresponding estimation results for 50 subsequent frames of this image sequence for one of the motion parameters (the rotation angle ϕ_x) in comparison to the data from the gyro-sensors. The dashed lines visualize the estimates derived from all extracted image point correspondences based on the minimization of the error functions F_0 and F_W. The solid lines represent the results after five iterations of the described detection of outliers. For the proposed conic error model a significant improvement of the accuracy (and the stability) is achieved. In contrast, the precision of the estimates derived from the objective function F_W is only little better than before. This indicates that the suggested half-perspective modeling restricted to object points in front of the camera is mandatory for a robust and accurate estimation of the ego-motion of the camera in case of moving 3D objects in the scene.

4.2.3 Comparison with Self-Calibration Methods

As Luong and Faugeras pointed out in [11], the best criterion to examine the applicability of a motion estimation algorithm is the stability and the precision of the computed epipoles, i.e. the corresponding projected image position of the estimated translational direction. This fact can be also deduced from our experiments with synthetic data presented in the previous subsection, since the absolute error of the translational component of the estimated motion dominates the overall error qualitatively and quantitatively.

Fig. 20 shows the Euclidean distances of the estimated epipoles from the projection of

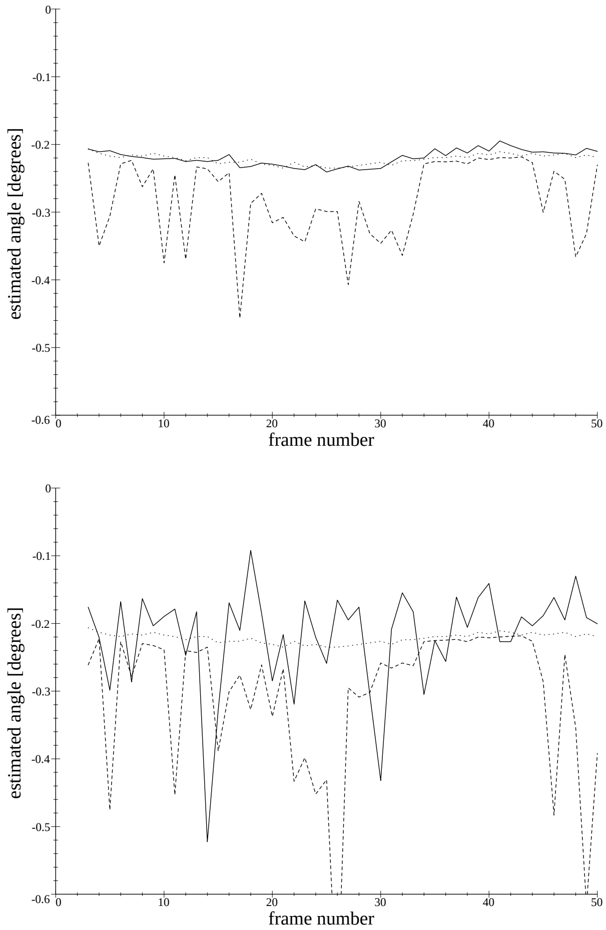

Figure 19: Estimation results for the angle ϕ_x based on the objective functions F_0 (top) and F_W (bottom); estimates without iterative refinement (dashed lines), estimates after detection of outliers (solid lines) and data from gyro-sensors (dotted lines)

the longitudinal axis of the test car for the calibrated approaches F_0 (based on the conic error model) and F_W (full-perspective modeling) in comparison with the results derived from the computation of the *Fundamental Matrix* and the objective function F_F in the uncalibrated case.

The deviations of the epipole positions achieved from the minimization of the objective functions F_0 and F_W amount to only 5 to 10 pixels. In contrast, the estimated positions of the epipoles derived from the *Fundamental Matrix* are scattered across the image with fluctuations of more than 50 pixels per frame. With the camera system used in our experiments, this corresponds to angular changes of the translational direction of 5° within 20 msec., which are absolutely incompatible with the motion of the car. Regarding the accurate and stable 3D reconstruction of distant objects, deviations of that magnitude are absolutely prohibitive: for instance in automotive applications, the angle between the viewing rays of a distant object point in consecutive images is by far smaller than one degree.

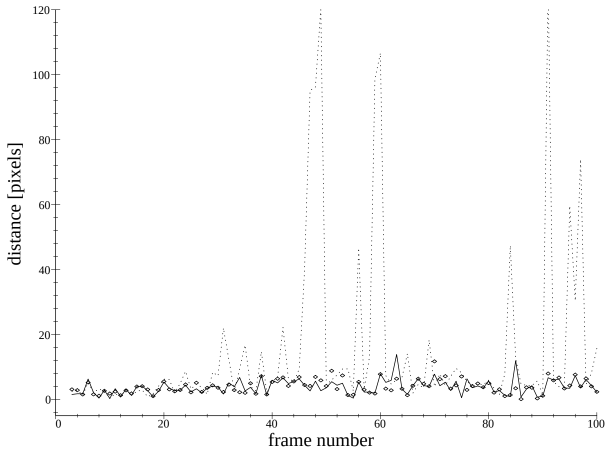

Figure 20: Euclidean distances of the estimated epipoles from the projected longitudinal axis of the test car; the solid lines and the dots visualize the results from F_0 and F_W, the dotted lines depict the differences for the estimates based on the computation of the Fundamental Matrix and the objective function F_F in the uncalibrated case

4.2.4 Influence of Simplifications, Optimal Solution versus Initial Values

In this section, the influence of the approximations introduced in section 3.5 is analyzed and discussed by a comparison of the results obtained from the simplified model (based on the minimization of the objective function F_5) to those derived from the full-scale approaches F_0 and F_W.

The results for the estimated rotational angles are presented in Fig. 21. Fig. 22 depicts the angular (horizontal and vertical) differences of the estimated translational directions from the longitudinal axis of the test car for the various approaches.

The figures show that the angular differences of the estimated translational direction derived from the simplified model as compared with the results of the full-scale approaches is typically smaller than $0.4°$. The corresponding rotational angles match even within $0.005°$.

From these experiments the following conclusions can be drawn: first, the simplified linear model is well-suited for the generation of initial values for the full-scale approach; second, since the initial values are already of high precision, the full-scale model F_0 converges towards the global minimum after only very few iterations; third, in special situations of automotive applications the accuracy achieved with the linear model is already sufficient and, in general, a suitable combination of the linear and the full-scale approach is well-prepared to meet the inherent requirements under real-time restrictions.

5 Discussion

The problem of computing the ego-motion of a camera from sets of point correspondences in a monocular image sequence has been analyzed. With the assumption of a calibrated camera, the so-called conic error model has been introduced. Taking advantage of this model, the contributions of distinct point correspondences to the objective function are treated perspectively invariant, i.e. independently of 3D distance. In addition and in

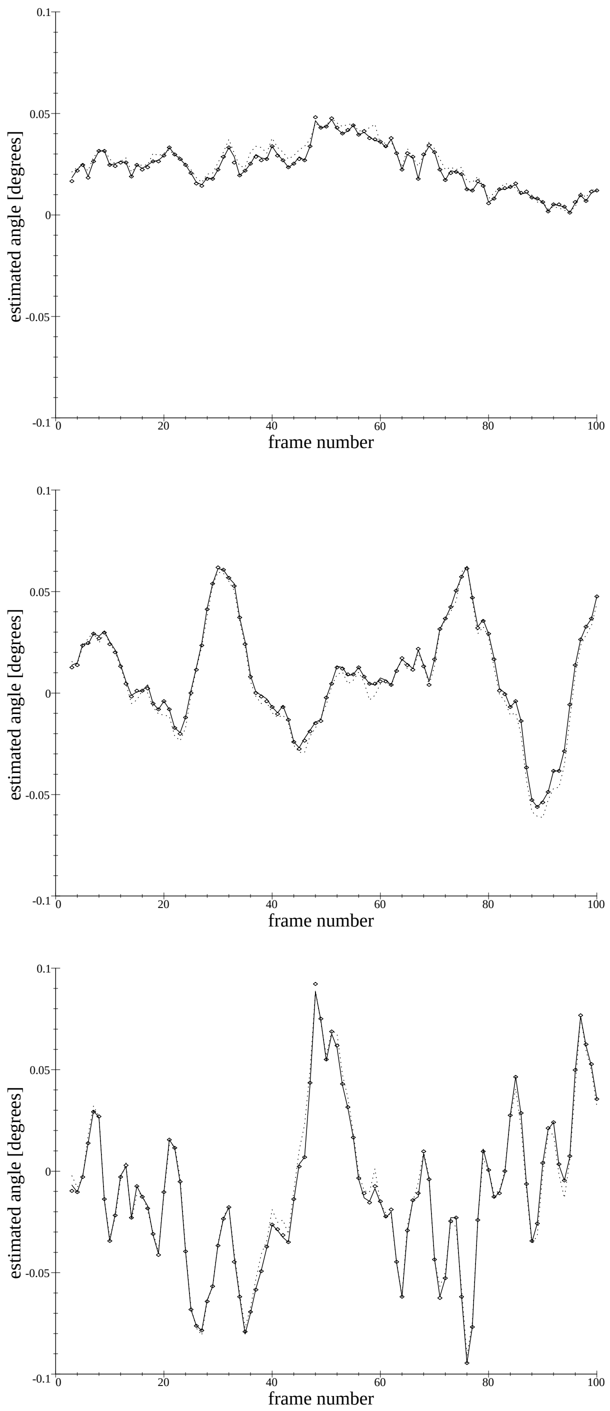

Figure 21: Estimates for the rotational motion parameters ϕ_x (top), ϕ_y (mid) and ϕ_z (bottom); the solid lines and the dots depict the results of the full-scale approaches F_0 and F_W, the dotted lines show the estimates of the simplified model F_5

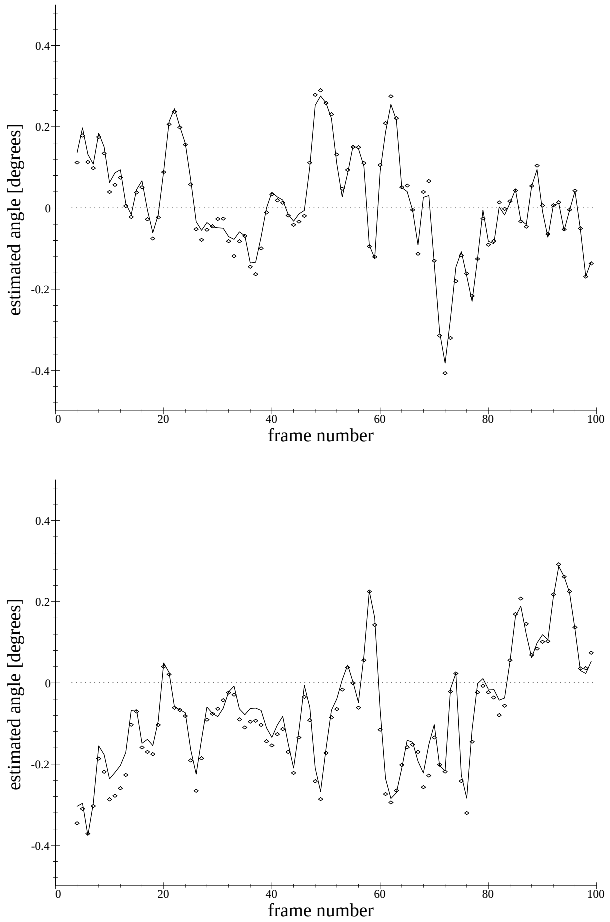

Figure 22: Angular horizontal (top) and vertical (bottom) differences of the estimates for t from the longitudinal axis of the test car; the solid lines and the dots depict the results of the full-scale approaches F_0 and F_W, the dotted lines show the estimates of the simplified model F_5

contrast to existing approaches in computer vision, this model permits to distinguish between feasible and non-feasible image correspondences related to 3D object points in front of and behind the camera, respectively.

Experiments proved that the algorithms based on this approach are highly accurate and stable. The minimization of the related objective function can be carried out in real-time. Therefore, and because of the capability to compute the ego-motion of a fast moving object with high precision, the approach is well-suited for automotive applications, e.g. vision based vehicle guidance.

The results presented in section 4 have been compared to recently published techniques based on a full-perspective modeling and the related epipolar geometry. It has been found that the proposed "half-perspective" point of view encourages the proper detection of mismatches and distinct relative motions for a given set of point correspondences. Furthermore, the stability of the achieved motion estimation with the described approach is significantly better than that of self-calibration methods based on the computation of the *Fundamental Matrix*.

As already outlined in the introduction, these properties are mainly ascribed to the decomposition of the global optimization task, which allows a more detailed and accurate formulation of individual error contributions. With the *Fundamental Matrix*, for instance, neither the integration of the geometric restriction to objects in front of the camera (inherent with the imaging of real optical systems), nor the proposed simplifications (e.g. decomposition of rotational and translational motion characteristics, affine approximation of the rotational components) applied in this paper would have been feasible. Experiments not presented here clearly show that the global optimization approach frequently suffers from interdependencies of model parameters that tend to minimize the global objective function but spoil the accuracy of individual motion parameters.

Corresponding to the overall robustness of the model presented in this paper, our experiments with synthetic as well as natural image data proved excellent accuracy of the estimation results.

Acknowledgment

We would like to thank the computer vision group at BMW for providing the animated and the natural image sequences and the car measurement data, including the results of camera calibration. We also acknowledge various helpful suggestions and comments concerning automotive applications by Dr. Robert Divko and Dr. Reiner Wertheimer. We are grateful to Dr. Wertheimer for his contribution to the final version of this manuscript and to Jürgen Haas at the University of Passau, who supplied the implementation to extract image point correspondences with sub-pixel precision.

A Proof of Theorem 1

We have to prove the respective values of

$$\begin{aligned}\varepsilon(r_1, r_2) &= \inf\{\psi \in [0,\pi] \mid C(r_1,\psi) \cap C(r_2,\psi) \neq \emptyset\}\\ &= \inf\{\psi \in [0,\pi] \mid C_{c_1,d_1,\cos(\psi)} \cap C_{c_2,d_2,\cos(\psi)} \neq \emptyset\}\end{aligned}$$

given in Theorem 1 for two viewing rays

$$\begin{aligned}r_1 &= \{m \in I\!R^3 \mid m = c_1 + \lambda d_1, \lambda \in I\!R, \lambda \geq 0\},\\ r_2 &= \{m \in I\!R^3 \mid m = c_2 + \sigma d_2, \sigma \in I\!R, \sigma \geq 0\},\end{aligned}$$

with a normalized displacement $c_2 - c_1$ and direction vectors $d_1, d_2 \in S_1(0)$.

Given two cones C_{c_1,d_1,α_1} and C_{c_2,d_2,α_2}, we say that C_{c_1,d_1,α_1} *intersects* C_{c_2,d_2,α_2} *perspectively*, if

$$C_{c_1,d_1,\alpha_1} \cap C_{c_2,d_2,\alpha_2} \neq \emptyset \quad \text{or} \quad C_{0,d_1,\alpha_1} \cap C_{0,d_2,\alpha_2} \neq \{0\}.$$

Since for every vector

$$e \in (C_{0,d_1,\alpha_1} \cap C_{0,d_2,\alpha_2}) \setminus \{0\},$$

the rays $\{c_1 + \lambda e \mid \lambda \geq 0\}$ and $\{c_2 + \lambda e \mid \lambda \geq 0\}$ are parallel and contained in C_{c_1,d_1,α_1} and C_{c_2,d_2,α_2}, respectively, the condition $C_{0,d_1,\alpha_1} \cap C_{0,d_2,\alpha_2} \neq \{0\}$ is equivalent to the existence of parallel rays in the cones C_{c_1,d_1,α_1} and C_{c_2,d_2,α_2}.

Vice versa, for parallel rays in the cones C_{c_1,d_1,α_1} and C_{c_2,d_2,α_2} with direction e, the inequality

$$\sphericalangle(d_1, d_2) \leq \sphericalangle(d_1, e) + \sphericalangle(e, d_2)$$

holds. In combination with the obvious relation

$$\varepsilon(r_1, r_2) \leq \frac{\sphericalangle(d_1, d_2)}{2} \leq \frac{\pi}{2},$$

this yields that the desired infimum can be rewritten as

$$\begin{aligned} \varepsilon(r_1, r_2) \;&=\; \inf\{\psi \in [0, \frac{\pi}{2}] \mid \\ &\qquad C_{c_1,d_1,\cos(\psi)} \text{ intersects } C_{c_2,d_2,\cos(\psi)} \text{ perspectively}\}. \end{aligned}$$

Moreover, the cos-function is strictly decreasing on $[0, \frac{\pi}{2}]$. Hence, we may consider the mapping

$$\kappa : (I\!R^3 \times S_1(0))^2 \to I\!R$$

defined by

$$\kappa(c_1, d_1, c_2, d_2) = \cos(\varepsilon(r_1, r_2))$$

rather than $\varepsilon(r_1, r_2)$. This leads to

$$\begin{aligned} \kappa(c_1, d_1, c_2, d_2) \;&=\; \sup\{\alpha \in [0, 1] | \\ &\qquad C_{c_1,d_1,\alpha} \text{ intersects } C_{c_2,d_2,\alpha} \text{ perspectively}\}. \end{aligned}$$

Finally, we may set $c_1 = 0$, because

$$x \in C_{c,d,\alpha} \Longleftrightarrow x - c \in C_{0,d,\alpha}.$$

With the preceding preparations, the proof of Theorem 1 can be formulated in terms of $\kappa(0, d_1, t, d_2)$ with $t = c_2 - c_1$ and $\|t\| = 1$. Starting with a lemma, which will be used repeatedly in the following arguments, the four cases addressed in Theorem 1 will be subsequently analyzed.

Lemma 1 *For any three vectors a, b and c with $\|c\| = 1$,*

$$\langle a|b\rangle \leq \langle a|c\rangle\langle b|c\rangle + \sqrt{\|a\|^2 - \langle a|c\rangle^2}\sqrt{\|b\|^2 - \langle b|c\rangle^2}.$$

Proof: Since the vector $a - \langle a|c\rangle c$ is orthogonal to c, we conclude from the Cauchy-Schwarz inequality

$$\begin{aligned} |\langle a|b\rangle - \langle a|c\rangle\langle b|c\rangle| \;&=\; |\langle a - \langle a|c\rangle c|b\rangle| \\ &=\; |\langle a - \langle a|c\rangle c|b - \langle b|c\rangle c\rangle| \\ &\leq\; \|a - \langle a|c\rangle c\| \; \|b - \langle b|c\rangle c\| \\ &=\; \sqrt{\|a\|^2 - \langle a|c\rangle^2}\sqrt{\|b\|^2 - \langle b|c\rangle^2}. \end{aligned}$$

■

Theorem 2 *For* $\alpha_0 = \frac{\|d_1+d_2\|}{2}$ *and* $\alpha \in \,]\alpha_0, 1]$*, the following relations hold:*

a) $C_{0,d_1,\alpha} \cap C_{0,d_2,\alpha} = \{0\}$.

b) If $\langle d_1 - d_2 | t\rangle \leq 0$*, then* $C_{0,d_1,\alpha} \cap C_{t,d_2,\alpha} = \emptyset$.

c) $d_1 + d_2 \in C_{0,d_1,\alpha_0} \cap C_{0,d_2,\alpha_0}$.

Proof: For $\alpha > \alpha_0$ and $x \in C_{0,d_1,\alpha} \setminus \{0\}$ the inequality

$$\langle d_1 - d_2 | x\rangle + \langle d_1 + d_2 | x\rangle = 2\langle d_1 | x\rangle = 2\alpha\ \|x\| > \|d_1 + d_2\|\ \|x\|$$

implies

$$\langle d_1 - d_2 | x\rangle > \|d_1 + d_2\|\ \|x\| - \langle d_1 + d_2 | x\rangle \geq 0. \tag{3}$$

Similarly, for $x \in C_{t,d_2,\alpha} \setminus \{t\}$

$$\langle d_1 - d_2 | x - t\rangle < 0. \tag{4}$$

a) According to (3) and (4), a point

$$x \in C_{0,d_1,\alpha} \cap C_{0,d_2,\alpha} \setminus \{0\}$$

had to satisfy the relation

$$\alpha\ \|x\| = \langle d_1 | x\rangle > \langle d_2 | x\rangle = \alpha\ \|x\|,$$

which is impossible.

b) For $x \in C_{0,d_1,\alpha} \setminus \{0\}$, (3) implies that

$$\langle d_1 - d_2 | x - \frac{t}{2}\rangle = \langle d_1 - d_2 | x\rangle - \frac{1}{2}\langle d_1 - d_2 | t\rangle > 0.$$

Similarly, (4) yields

$$\langle d_1 - d_2 | x - \frac{t}{2}\rangle < 0,$$

for $x \in C_{t,d_2,\alpha} \setminus \{t\}$. This excludes the existence of a common point in both cones.

c) Since

$$\begin{aligned}\langle d_1 | d_1 + d_2\rangle &= 1 + \langle d_1 | d_2\rangle = \frac{\|d_1 + d_2\|}{2}\|d_1 + d_2\| = \alpha_0 \|d_1 + d_2\| \\ &= \langle d_2 | d_1 + d_2\rangle,\end{aligned}$$

$d_1 + d_2 \in C_{0,d_1,\alpha_0} \cap C_{0,d_2,\alpha_0}$.

■

Projecting the vectors d_1 and d_2 onto the plane orthogonal to t yields the vectors

$$\begin{aligned}\tilde{d}_1 &= d_1 - \langle d_1|t\rangle t,\\ \tilde{d}_2 &= d_2 - \langle d_2|t\rangle t,\end{aligned}$$

which are used repeatedly in the following arguments. The norm and the inner product of these vectors are given by

$$\begin{aligned}\|\tilde{d}_1\|^2 &= 1 - \langle d_1|t\rangle^2,\\ \|\tilde{d}_2\|^2 &= 1 - \langle d_2|t\rangle^2,\\ \langle \tilde{d}_1|\tilde{d}_2\rangle &= \langle d_1|d_2\rangle - \langle d_1|t\rangle\langle d_2|t\rangle.\end{aligned}$$

Theorem 3 *If $\langle d_1 - d_2|t\rangle > 0$ and $\|\tilde{d}_1\|^2 + \langle \tilde{d}_1|\tilde{d}_2\rangle \le 0$, the following relations hold for $\alpha_1 = \langle d_1|t\rangle$:*

a) If $\alpha \in]\alpha_1, 1]$, then $C_{0,d_1,\alpha} \cap C_{0,d_2,\alpha} = \{0\}$ and $C_{0,d_1,\alpha} \cap C_{t,d_2,\alpha} = \emptyset$.

b) $t \in C_{0,d_1,\alpha_1} \cap C_{t,d_2,\alpha_1}$.

Proof: Suppose $\langle d_1|t\rangle \le 0$. Then,

$$\begin{aligned}\|\tilde{d}_2\|^2 + \langle \tilde{d}_1|\tilde{d}_2\rangle &= 1 - \langle d_2|t\rangle^2 + \langle \tilde{d}_1|\tilde{d}_2\rangle\\ &< 1 - \langle d_1|t\rangle^2 + \langle \tilde{d}_1|\tilde{d}_2\rangle\\ &= \|\tilde{d}_1\|^2 + \langle \tilde{d}_1|\tilde{d}_2\rangle \le 0.\end{aligned}$$

and, therefore,

$$\|\tilde{d}_1 + \tilde{d}_2\|^2 = \|\tilde{d}_1\|^2 + \|\tilde{d}_2\|^2 + 2\langle \tilde{d}_1|\tilde{d}_2\rangle < 0,$$

which is impossible. Hence, $\alpha_1 = \langle d_1|t\rangle > 0$.

a) Because

$$\begin{aligned}0 &\ge \|\tilde{d}_1\|^2 + \langle \tilde{d}_1|\tilde{d}_2\rangle\\ &= 1 - \langle d_1|t\rangle^2 + \langle d_1|d_2\rangle - \langle d_1|t\rangle\langle d_2|t\rangle\\ &= \frac{\|d_1 + d_2\|^2}{2} - \left(\langle d_1|t\rangle^2 + \langle d_1|t\rangle\langle d_2|t\rangle\right),\end{aligned}$$

it follows that

$$\frac{\|d_1 + d_2\|^2}{2} \le \langle d_1|t\rangle^2 + \langle d_1|t\rangle\langle d_2|t\rangle < 2\langle d_1|t\rangle^2,$$

and, therefore, $\alpha_1 > \alpha_0$. From Theorem 2.a), we conclude that

$$C_{0,d_1,\alpha} \cap C_{0,d_2,\alpha} = \{0\}.$$

Furthermore, note that

$$\|\tilde{d}_1\|^2 = 1 - \langle d_1|t\rangle^2 > 1 - \alpha^2,$$

which yields $\tilde{d}_1 \ne 0$,

$$\alpha_1 = \sqrt{1 - \|\tilde{d}_1\|^2}, \quad \text{and} \quad \|\tilde{d}_1\| > \sqrt{1 - \alpha^2}.$$

Moreover, since $\|\tilde{d}_1\|^2 + \langle \tilde{d}_1 | \tilde{d}_2 \rangle \leq 0$,

$$\langle \tilde{d}_1 | \tilde{d}_2 \rangle \leq -\|\tilde{d}_1\|^2, \quad \text{and} \quad \|\tilde{d}_1\|^4 \leq \langle \tilde{d}_1 | \tilde{d}_2 \rangle^2.$$

Because $\tilde{d}_1$ is orthogonal to t, Lemma 1 yields for $x \in C_{t,d_2,\alpha} \setminus \{t\}$,

$$\begin{aligned}
\langle \tilde{d}_1 | x \rangle &= \langle \tilde{d}_1 | x - t \rangle \\
&\leq \langle \tilde{d}_1 | d_2 \rangle \langle x - t | d_2 \rangle + \sqrt{\|\tilde{d}_1\|^2 - \langle \tilde{d}_1 | d_2 \rangle^2} \sqrt{\|x - t\|^2 - \langle x - t | d_2 \rangle^2} \\
&= \langle \tilde{d}_1 | \tilde{d}_2 \rangle \; \alpha \|x - t\| + \sqrt{\|\tilde{d}_1\|^2 - \langle \tilde{d}_1 | \tilde{d}_2 \rangle^2} \sqrt{\|x - t\|^2 (1 - \alpha^2)} \\
&\leq \|x - t\| \left(-\|\tilde{d}_1\|^2 \alpha + \|\tilde{d}_1\| \sqrt{\|\tilde{d}_1\|^2 - \|\tilde{d}_1\|^4} \right) \\
&\leq \|x - t\| \left(-\|\tilde{d}_1\|^2 \alpha + \|\tilde{d}_1\|^2 \alpha_1 \right) \\
&= \|\tilde{d}_1\|^2 \|x - t\| \, (\alpha_1 - \alpha) < 0.
\end{aligned}$$

On the other hand, for $x \in C_{0,d_1,\alpha} \setminus \{0\}$,

$$\begin{aligned}
\langle \tilde{d}_1 | x \rangle &= \langle d_1 | x \rangle - \langle d_1 | t \rangle \langle t | x \rangle \\
&= \alpha \; \|x\| - \alpha_1 \langle t | x \rangle \\
&> \alpha_1 \; (\|x\| - \langle t | x \rangle) \geq 0.
\end{aligned}$$

This excludes the existence of a common point in both cones.

b) The proof is trivial, since $\langle d_1 | t \rangle = \alpha_1$.

■

Theorem 4 *If $\langle d_1 - d_2 | t \rangle > 0$ and $\|\tilde{d}_2\|^2 + \langle \tilde{d}_1 | \tilde{d}_2 \rangle \leq 0$, the following relations hold for $\alpha_2 = -\langle d_2 | t \rangle$:*

a) If $\alpha \in \,]\alpha_2, 1]$, then $C_{0,d_1,\alpha} \cap C_{0,d_2,\alpha} = \{0\}$ and $C_{0,d_1,\alpha} \cap C_{t,d_2,\alpha} = \emptyset$.

b) $0 \in C_{0,d_1,\alpha_2} \cap C_{t,d_2,\alpha_2}$.

Proof: Similar to Theorem 3. ■

Lemma 2 *a) $\langle (d_1 + d_2) \times t | d_2 \rangle = \langle d_1 \times t | d_2 \rangle$.*

b) $\langle (d_1 + d_2) \times t | d_2 \rangle = -\langle (d_1 + d_2) \times t | d_1 \rangle$.

c) $\|\tilde{d}_1 + \tilde{d}_2\|^2 = \|(d_1 + d_2) \times t\|^2$.

d) $\|d_1 - d_2\|^2 \|(d_1 + d_2) \times t\|^2 \geq 4 \langle d_1 \times t | d_2 \rangle^2$.

Proof:

a) Since $\langle d_2 \times t | d_2 \rangle = 0$, this equality follows from

$$\langle (d_1 + d_2) \times t | d_2 \rangle = \langle d_1 \times t | d_2 \rangle + \langle d_2 \times t | d_2 \rangle.$$

b) This equation is proved by

$$\begin{aligned} 0 &= \langle (d_1 + d_2) \times t | d_1 + d_2 \rangle \\ &= \langle (d_1 + d_2) \times t | d_1 \rangle + \langle (d_1 + d_2) \times t | d_2 \rangle. \end{aligned}$$

c) From the definition of $\tilde{d}_1$ and $\tilde{d}_2$, it follows that

$$\begin{aligned} \|\tilde{d}_1 + \tilde{d}_2\|^2 &= \|\tilde{d}_1\|^2 + \|\tilde{d}_2\|^2 + 2\langle \tilde{d}_1 | \tilde{d}_2 \rangle \\ &= 2 - \langle d_1 | t \rangle^2 - \langle d_2 | t \rangle^2 + 2\langle d_1 | d_2 \rangle - 2\langle d_1 | t \rangle \langle d_2 | t \rangle \\ &= \|(d_1 + d_2)\|^2 \|t\|^2 - \langle d_1 + d_2 | t \rangle^2 \\ &= \|(d_1 + d_2) \times t\|^2. \end{aligned}$$

d) Since

$$\begin{aligned} \langle d_1 \times t | d_2 \rangle^2 &= \begin{vmatrix} \langle d_1 | d_1 \rangle & \langle d_1 | t \rangle & \langle d_1 | d_2 \rangle \\ \langle t | d_1 \rangle & \langle t | t \rangle & \langle t | d_2 \rangle \\ \langle d_2 | d_1 \rangle & \langle d_2 | t \rangle & \langle d_2 | d_2 \rangle \end{vmatrix} \\ &= 1 - \langle d_1 | t \rangle^2 - \langle d_2 | t \rangle^2 - \langle d_1 | d_2 \rangle^2 + 2\langle d_1 | d_2 \rangle \langle d_1 | t \rangle \langle d_2 | t \rangle, \end{aligned}$$

we conclude that

$$\begin{aligned} &\|d_1 - d_2\|^2 \|(d_1 + d_2) \times t\|^2 - 4\langle d_1 \times t | d_2 \rangle^2 \\ &= 2(1 - \langle d_1 | d_2 \rangle) \left[2 + 2\langle d_1 | d_2 \rangle - \langle d_1 | t \rangle^2 - \langle d_2 | t \rangle^2 - 2\langle d_1 | t \rangle \langle d_2 | t \rangle \right] \\ &\quad -4 \left[1 - \langle d_1 | t \rangle^2 - \langle d_2 | t \rangle^2 - \langle d_1 | d_2 \rangle^2 + 2\langle d_1 | d_2 \rangle \langle d_1 | t \rangle \langle d_2 | t \rangle \right] \\ &= 2\langle d_1 | t \rangle^2 + 2\langle d_2 | t \rangle^2 - 4\langle d_1 | t \rangle \langle d_2 | t \rangle \\ &\quad +\langle d_1 | d_2 \rangle \left[2\langle d_1 | t \rangle^2 + 2\langle d_2 | t \rangle^2 - 4\langle d_1 | t \rangle \langle d_2 | t \rangle \right] \\ &= 2(1 + \langle d_1 | d_2 \rangle)(\langle d_1 | t \rangle - \langle d_2 | t \rangle)^2 \geq 0. \end{aligned}$$

■

Lemma 3 *For $d_1 + d_2 \neq 0$ and $\langle d_1 | t \rangle \neq \langle d_2 | t \rangle$ the following relations hold:*

a) The three vectors

$$\begin{aligned} n_1 &= t, \\ n_2 &= \frac{(d_1 + d_2) \times t}{\|(d_1 + d_2) \times t\|}, \\ n_3 &= \frac{\tilde{d}_1 + \tilde{d}_2}{\|\tilde{d}_1 + \tilde{d}_2\|}, \end{aligned}$$

are well-defined and orthonormal in $\mathbb{R}^3$.

b) For

$$\begin{aligned} \xi &= (1 + \langle d_1 | d_2 \rangle) \langle d_1 - d_2 | t \rangle, \\ \lambda &= \frac{\|\tilde{d}_2\|^2 + \langle \tilde{d}_1 | \tilde{d}_2 \rangle}{\xi}, \\ \mu &= \frac{\|\tilde{d}_1\|^2 + \langle \tilde{d}_1 | \tilde{d}_2 \rangle}{\xi}, \end{aligned}$$

and

$$\begin{aligned} m_1 &= \lambda(d_1 - \langle n_2|d_1\rangle n_2), \\ m_2 &= \mu(d_2 - \langle n_2|d_2\rangle n_2) + t, \end{aligned}$$

the vectors m_1 and m_2 are identical.

Proof:

a) Suppose $(d_1 + d_2) \times t = 0$. Then $\sigma(d_1 + d_2) = t$ for some $\sigma \in \mathbb{R}$, $\sigma \neq 0$, since $t \neq 0$ and $d_1 + d_2 \neq 0$. This leads to $\langle d_1 - d_2|t\rangle = \sigma\langle d_1 - d_2|d_1 + d_2\rangle = 0$, which contradicts $\langle d_1|t\rangle \neq \langle d_2|t\rangle$. Hence, $(d_1 + d_2) \times t \neq 0$ and Lemma 2.c) yields $\tilde{d}_1 + \tilde{d}_2 \neq 0$.

Therefore, the vectors n_1, n_2 and n_3 are well-defined and by normalization their Euclidean lengths $\|n_1\|$, $\|n_2\|$ and $\|n_3\|$ are equal to 1. $\langle n_1|n_2\rangle = 0$, since $\langle (d_1 + d_2) \times t|t\rangle = 0$. $\langle n_1|n_3\rangle = 0$, because $\langle \tilde{d}_1|t\rangle = 0$ and $\langle \tilde{d}_2|t\rangle = 0$. Due to the equality

$$t \times ((d_1 + d_2) \times t) = (d_1 + d_2) - \langle d_1 + d_2|t\rangle t = \tilde{d}_1 + \tilde{d}_2,$$

n_2 is orthogonal to $\tilde{d}_1 + \tilde{d}_2$ and, therefore $\langle n_2|n_3\rangle = 0$.

b) The inner products of m_1 and m_2 with the vector n_1 are given by

$$\begin{aligned} \langle m_1|n_1\rangle &= \lambda\langle d_1|n_1\rangle, \\ \langle m_2|n_1\rangle &= \mu\langle d_2|n_1\rangle + 1. \end{aligned}$$

Since $\xi \neq 0$ and

$$\begin{aligned} &\xi(\langle m_1|n_1\rangle - \langle m_2|n_1\rangle) = \\ &\quad= (\|\tilde{d}_2\|^2 + \langle \tilde{d}_1|\tilde{d}_2\rangle)\langle d_1|t\rangle - (\|\tilde{d}_1\|^2 + \langle \tilde{d}_1|\tilde{d}_2\rangle)\langle d_2|t\rangle - \xi \\ &\quad= \left[1 - \langle d_2|t\rangle^2 + \langle d_1|d_2\rangle - \langle d_1|t\rangle\langle d_2|t\rangle\right]\langle d_1|t\rangle \\ &\quad\quad - \left[1 - \langle d_1|t\rangle^2 + \langle d_1|d_2\rangle - \langle d_1|t\rangle\langle d_2|t\rangle\right]\langle d_2|t\rangle - \xi \\ &\quad= \langle d_1|t\rangle + \langle d_1|d_2\rangle\langle d_1|t\rangle - \langle d_2|t\rangle - \langle d_1|d_2\rangle\langle d_2|t\rangle - \xi \\ &\quad= \xi - \xi = 0, \end{aligned}$$

the inner products $\langle m_1|n_1\rangle$ and $\langle m_2|n_1\rangle$ are identical.

Because n_2 is orthogonal to t and $\|n_2\| = 1$,

$$\begin{aligned} \langle m_1|n_2\rangle &= \lambda(\langle d_1|n_2\rangle - \langle n_2|d_1\rangle) = 0, \\ \langle m_2|n_2\rangle &= \mu(\langle d_2|n_2\rangle - \langle n_2|d_2\rangle) = 0. \end{aligned}$$

Since $t = n_1$,

$$\|\tilde{d}_1\|^2 + \langle \tilde{d}_1|\tilde{d}_2\rangle = \langle \tilde{d}_1 + \tilde{d}_2|\tilde{d}_1\rangle = \langle \tilde{d}_1 + \tilde{d}_2|d_1\rangle,$$

and, similarly,

$$\|\tilde{d}_2\|^2 + \langle \tilde{d}_1|\tilde{d}_2\rangle = \langle \tilde{d}_1 + \tilde{d}_2|d_2\rangle,$$

the inner products

$$\langle m_1|n_3\rangle = \lambda\langle d_1|n_3\rangle = \frac{\langle \tilde{d}_1+\tilde{d}_2|d_2\rangle\langle \tilde{d}_1+\tilde{d}_2|d_1\rangle}{\xi\ \|\tilde{d}_1+\tilde{d}_2\|},$$
$$\langle m_2|n_3\rangle = \mu\langle d_2|n_3\rangle = \frac{\langle \tilde{d}_1+\tilde{d}_2|d_1\rangle\langle \tilde{d}_1+\tilde{d}_2|d_2\rangle}{\xi\ \|\tilde{d}_1+\tilde{d}_2\|},$$

are identical, too.

Hence, all coordinates of m_1 and m_2 with respect to the orthonormal basis $\{n_1, n_2, n_3\}$ are identical. Therefore, $m_1 = m_2$.

■

For the subsequent arguments, we define

$$n = n_2 = \frac{(d_1+d_2)\times t}{\|(d_1+d_2)\times t\|}.$$

Theorem 5 *If* $\langle d_1 - d_2|t\rangle > 0$ *and*

$$\|\tilde{d}_1\|^2 + \langle \tilde{d}_1|\tilde{d}_2\rangle \quad > \quad 0, \tag{5}$$
$$\|\tilde{d}_2\|^2 + \langle \tilde{d}_1|\tilde{d}_2\rangle \quad > \quad 0, \tag{6}$$

the following relations hold for

$$\alpha_3 = \sqrt{1-\langle n|d_2\rangle^2}:$$

a) If $\alpha \in\,]\alpha_3, 1]$, *then* $C_{0,d_1,\alpha}\cap C_{0,d_2,\alpha} = \{0\}$ *and* $C_{0,d_1,\alpha}\cap C_{t,d_2,\alpha} = \emptyset$.

b) $m_1 = m_2 \in C_{0,d_1,\alpha_3}\cap C_{t,d_2,\alpha_3}$.

Proof: Note that the requirements of Lemma 3 are fulfilled, since $\langle d_1|t\rangle > \langle d_2|t\rangle$ and (5) and (6) imply $d_1 + d_2 \neq 0$.

a) From Lemma 2.d) we have

$$\frac{\|d_1-d_2\|^2}{4} \geq \frac{\langle d_1\times t|d_2\rangle^2}{\|(d_1+d_2)\times t\|^2} = \langle n|d_2\rangle^2.$$

Thus,

$$\frac{\|d_1+d_2\|^2}{4} = 1 - \frac{\|d_1-d_2\|^2}{4} \leq 1-\langle n|d_2\rangle^2 = \alpha_3^2,$$

which yields $\alpha_3 \geq \alpha_0$, and, therefore, by Theorem 2.a)

$$C_{0,d_1,\alpha}\cap C_{0,d_2,\alpha} = \{0\}.$$

For $x \in C_{t,d_2,\alpha}\setminus\{t\}$, we conclude from Lemma 1, since t is orthogonal to n and $|\langle n|d_2\rangle| > \sqrt{1-\alpha^2}$,

$$\begin{aligned}\langle n|x\rangle &= \langle n|x-t\rangle \\ &\leq \langle n|d_2\rangle\langle x-t|d_2\rangle + \sqrt{1-\langle n|d_2\rangle^2}\sqrt{\|x-t\|^2-\langle x-t|d_2\rangle^2} \\ &= \langle n|d_2\rangle\alpha\|x-t\| + \alpha_3\|x-t\|\sqrt{1-\alpha^2} \\ &\leq \alpha\|x-t\|(\langle n|d_2\rangle + \sqrt{1-\alpha^2}) \\ &< \alpha\|x-t\|(\langle n|d_2\rangle + |\langle n|d_2\rangle|),\end{aligned} \tag{7}$$

and, analogously,

$$-\langle n|x\rangle < \alpha\|x-t\|(-\langle n|d_2\rangle + |\langle n|d_2\rangle|). \tag{8}$$

Likewise, for $x \in C_{0,d_1,\alpha} \setminus \{0\}$, we get

$$\begin{aligned} \langle n|x\rangle &< \alpha\|x\|(\langle n|d_1\rangle + |\langle n|d_1\rangle|), && (9)\\ -\langle n|x\rangle &< \alpha\|x\|(-\langle n|d_1\rangle + |\langle n|d_1\rangle|). && (10)\end{aligned}$$

Because $\langle n|d_1\rangle = -\langle n|d_2\rangle$ (Lemma 2.b)), this implies that for any

$$x \in C_{0,d_1,\alpha} \cap C_{t,d_2,\alpha} \setminus \{0,t\},$$

$\langle n|x\rangle < 0$ and $-\langle n|x\rangle < 0$, which is impossible.
Because $\langle n|0\rangle = 0$ and $t \neq 0$, (7) and (8) yield

$$0 \notin C_{t,d_2,\alpha} \subset C_{0,d_1,\alpha} \cap C_{t,d_2,\alpha}.$$

Because $\langle n|t\rangle = 0$ and $t \neq 0$, (9) and (10) yield

$$t \notin C_{0,d_1,\alpha} \subset C_{0,d_1,\alpha} \cap C_{t,d_2,\alpha}.$$

Hence the intersection of the cones is empty.

b) Since

$$\|m_1\|^2 = \lambda^2\|d_1 - \langle n|d_1\rangle n\|^2 = \lambda^2(1-\langle n|d_1\rangle^2) = \lambda^2\alpha_3^2$$

and $\lambda > 0$, $\|m_1\| = \lambda\alpha_3$. This leads to

$$\langle d_1|m_1\rangle = \lambda(1-\langle n|d_1\rangle^2) = \lambda\alpha_3^2 = \alpha_3\|m_1\|,$$

which yields $m_1 \in C_{0,d_1,\alpha_3}$. Similarly, because

$$\|m_2 - t\|^2 = \mu^2\|d_2 - \langle n|d_2\rangle n\|^2 = \mu^2\alpha_3^2.$$

and $\mu > 0$, $\|m_2 - t\| = \mu\alpha_3$. This leads to

$$\langle d_2|m_2 - t\rangle = \mu(1-\langle n|d_2\rangle^2) = \mu\alpha_3^2 = \alpha_3\|m_2 - t\|,$$

which yields $m_2 \in C_{t,d_2,\alpha_3}$. From Lemma 3 we conclude that

$$m_1 = m_2 \in C_{0,d_1,\alpha_3} \cap C_{t,d_2,\alpha_3}.$$

■

Theorem 6 *a) If* $\langle d_1 - d_2 | t\rangle \leq 0$*, then* $\kappa(0, d_1, t, d_2) = \alpha_0$.

b) If $\langle d_1 - d_2 | t\rangle > 0$ *and* $\|\tilde{d}_1\|^2 + \langle \tilde{d}_1 | \tilde{d}_2\rangle \leq 0$*, then* $\kappa(0, d_1, t, d_2) = \alpha_1$.

c) If $\langle d_1 - d_2 | t\rangle > 0$ *and* $\|\tilde{d}_2\|^2 + \langle \tilde{d}_1 | \tilde{d}_2\rangle \leq 0$*, then* $\kappa(0, d_1, t, d_2) = \alpha_2$.

d) If $\langle d_1 - d_2 | t\rangle > 0$, $\|\tilde{d}_1\|^2 + \langle \tilde{d}_1 | \tilde{d}_2\rangle > 0$ *and* $\|\tilde{d}_2\|^2 + \langle \tilde{d}_1 | \tilde{d}_2\rangle > 0$*, then* $\kappa(0, d_1, t, d_2) = \alpha_3$.

Proof:

a) From Theorem 2, the cones $C_{0,d_1,\alpha}$ and $C_{t,d_2,\alpha}$ do not intersect perspectively for $\alpha > \alpha_0$. Since $d_1 + d_2 \in C_{0,d_1,\alpha_0} \cap C_{0,d_2,\alpha_0}$, the desired supremum is given by α_0 for $d_1 + d_2 \neq 0$. For $d_1 + d_2 = 0$,

$$C_{0,d_1,\alpha_0} \cap C_{0,d_2,\alpha_0} = C_{0,d_1,0} = \{x \in \mathbb{R}^3 \mid \langle d_1 | x\rangle = 0\} \supset \{0\}.$$

Hence, $\kappa(0, d_1, t, d_2) = \alpha_0$.

b)-d) Analogously from Theorem 3, Theorem 4 and Theorem 5.

■

References

1. J.W. Roach and J.K. Aggarwal, Determing the Movement of Objects from a Sequence of Images, *IEEE Trans. Pattern Analysis and Machine Intelligence* **2**, No. 6, 1980, 554–562.

2. R.Y. Tsai and T.S. Huang, Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces, *IEEE Trans. Pattern Analysis and Machine Intelligence* **6**, No. 1, 1984, 13–27.

3. H.C. Longuet-Higgins, A computer algorithm for reconstructing a scene from two projections, *Nature* **293**, 1981, 133–135.

4. M.E. Spetsakis and J. Aloimonos, Optimal computing of structure from motion using point correspondences, in *Proceedings, 2nd International Conference on Computer Vision, Tampa, USA, December 1988*, pp. 449–453.

5. J. Weng, T.S. Huang and N. Ahuja, *Motion and Structure from Image Sequences*, Springer-Verlag, Berlin, 1993.

6. J. Weng, T.S. Huang and N. Ahuja, Optimal Motion and Structure Estimation, *IEEE Trans. Pattern Analysis and Machine Intelligence* **15**, No. 9, 1993, 864–884.

7. J. Philip, Estimation of Three-Dimensional Motion of Rigid Objects from Noisy Observations, *IEEE Trans. Pattern Analysis and Machine Intelligence* **13**, No. 1, 1991, 61–66.

8. T.S. Huang and A.N. Netravali, Motion and Structure from Feature Correspondences: A Review, *Proc. IEEE* **82**, No. 2, 1994, 252–268.

9. S.J. Maybank, *Theory of Reconstruction from Image Motion*, Springer-Verlag, Berlin, 1993.

10. O.D. Faugeras, Q.T. Luong and S.J. Maybank, Camera Self-Calibration: Theory and Experiments, in *Proceedings, 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy, May 1992*, pp. 321–334.

11. Q.T. Luong and O.D. Faugeras, The Fundamental Matrix: Theory, Algorithms and Stability Analysis, *International Journal Computer Vision* **17**, 1996, 43–75.

12. O.D. Faugeras and L. Robert, What Can Two Images Tell Us About a Third One?, *International Journal Computer Vision* **18**, 1996, 5–19.

13. T. Viéville, O.D. Faugeras and Q.T. Luong, Motion of Points and Lines in the Uncalibrated Case, *International Journal Computer Vision* **17**, 1996, 7–41.

14. R.Y. Tsai, A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses, *IEEE Journal of Robotics and Automation* **3**, No. 4, 1987, 323–344.

15. Z. Zhang, On the Epipolar Geometry Between Two Images With Lens Distortion, in *Proceedings, 13th International Conference on Pattern Recognition, Vienna, Austria, August 1996*, pp. 407–411.