# Active Intrinsic Calibration using Vanishing Points 

Konstantinos Daniilidis and Jörg Ernst<br>Computer Science Institute, Kiel University


#### Abstract

Abstact During a camera motion with arbitrary translation and fixed axis rotation a vanishing point traverses a conic section on the image plane. The intrinsic parameters of the camera can be obtained from the coefficients of the conic. Special care is given to the propagation of the error covariances and the elimination of the statistical bias in conic fitting.


Keywords: Calibration, active vision, vanishing points, statistical bias

## Address for correspondence:

Konstantinos Daniilidis
Computer Science Institute, Kiel University
Preusserstr. 1-9, 24105 Kiel, Germany
FAX: +49 431560481 , Tel: +49 431560485
email: kdinformatik.uni-kiel.de

## 1 Introduction

The intrinsic parameters of a camera fix the relation between image pixel coordinates and ray directions in the 3D camera coordinate system. In the nineties many approaches (see (Faugeras, 1992)) were proposed that solve visual tasks without knowledge of the intrinsic parameters. However, any visual capability involving metric information necessitates the mapping between image plane and ray directions. The transition between absent and complete knowledge of the intrinsic parameters has been formalized in geometric terms in the exposition (Faugeras, 1995).

We are interested here in calibration procedures suitable for active camera systems mounted on mobile robots. The use of the degrees of freedom of the camera introduces additional information compared to static camera systems. However, the use of an active camera on a mobile robot poses some requirements which might not be met by traditional calibration as used for 3D-measurement tasks in a static environment. This implies the need to get rid of special calibration objects like plates and cubes with known geometry in world coordinates. Second, the camera should be able to be calibrated without interruption of the robot's motion. Third, calibration should use image features that allow accurate image position estimation and robust matching among multiple views. Fourth, calibration of the intrinsic parameters should be decoupled from the computation of the extrinsic parameters which relate a fixed world coordinate system with the camera coordinate system. Of course, we do not expect to achieve the same precision as well established techniques like Tsai's (1986) which use hundreds of points with exactly known world coordinates as well as data from the camera specification. According to the described requirements we present in this paper a new approach for the computation of the intrinsic parameters making following two assumptions:

1. The only assumption on the world is the existence of one set of parallel lines.
2. The active camera should be able to rotate around the two axes of the camera coordinate system (tilt and pan).

Lines are image features with parameters that can be estimated robustly. Hough transform techniques exist (Lutton et al., 1994; Palmer et al., 1993) that group lines into sets of projections of parallel lines. If the line direction is not parallel to the image plane the projections of parallel lines intersect at the vanishing point. When a camera is arbitrarily moving the vanishing points change their position only due to camera rotation. In practice, large rotation angles are required to obtain accurate estimates of the intrinsic parameters. The family of parallel lines should thus subtend a wide visual angle. However, this requirement can be attenuated by allowing the camera to translate since translation does not affect the vanishing points. A further advantage regarding active systems is that intrinsic calibration can be carried out before sensor-effector calibration since a deviation of the projection center from the rotation axes introduces just a translation. However, the camera axes should be adjusted to be parallel to the rotation axes of the mounting platform.

If the rotation axis remains fixed the projection of the trajectory of a point is a conic section. The coefficients of the conic section contain information on the axis and the intrinsic parameters. We illustrate in Fig. (1) such a hyperbola in the image plane produced by a rotation of the camera around the vertical axis. If the rotation is around
the $x$ - or $y$ - axis the solution for the intrinsic parameters is in closed form but only a subset of the parameters can be computed from every conic section. Using both rotations we are able to obtain all the intrinsic parameters in closed form.


Figure 1. A rotation of the camera around the vertical $Y_{s}$-axis is equivalent to the rotation of a point about this axis. The circular cone including the point trajectory intersects the image plane in a hyperbola.

No iterative minimization is needed in any of the algorithmic steps. We put special emphasis on the propagation of the error through covariances in the intermediate estimates. Using the methodology proposed by Kanatani (1993) we eliminate the bias in the computation of the coefficients of the conic section.

We will show in the experiments that the image centers can be computed more reliably than the scaling factors. This fact can be geometrically interpreted by means of the point distribution constraining the conic to be fitted which in our case will be shown to be a hyperbola. The recovery of the image center depends on the feasibility of the symmetry axes estimation whereas the recovery of the scaling factors depends on the slopes of the asymptotes. In fact, the limited span of the point distribution allows only the reliable
estimation of the hyperbola symmetry axis but not of the asymptotes, hence the more accurate estimation of the image centers.

The main factor affecting accuracy is the amount of rotation carried out by the camera. We test the sensitivity of the approach to error in the line parameters. We compare our results to results obtained by the implementation of another active calibration approach that uses point trajectories under varying zoom to compute the image center and orthogonal sets of parallel lines (conjugate vanishing points) to estimate the scaling factors (Li, 1994).

We begin with a review of approaches which avoid using the world coordinates of calibration points. We continue with the mathematical description of the algorithm. Then we delve into the statistics of conic fitting. We finish by describing our experimental results both on synthetic as well as real data.

## 2 Related Work

Considering related work we will refer only to approaches that use motion in order to calibrate the intrinsic parameters without knowledge of world points or angles. We just mention that a couple of approaches (Caprile \& Torre, 1990; Echigo, 1990) exist using vanishing points and stationary cameras but they assume knowledge of the angles between the representative parallels. Neither of them can compute all the intrinsics parameters.

Since the seminal paper by Maybank and Faugeras (1992) it is known that intrinsic parameters encoded in the image of the absolute conic can be obtained from point correspondences in four views. Later it was proven (Luong \& Vieville, 1994) that the image of the absolute conic can also be obtained from point correspondences of three views if they are projections from points on the plane at infinity. For projections of points with finite depth this is equivalent to three views arising from two pure rotations of the camera (Hartley, 1994). Although the computation of the intrinsics is in principle linear according to the latter approaches a non-linear refinement step is applied to increase accuracy. Dron's work (1993) can be classified in the same framework of absent or known translation, however, it uses only linear least squares techniques.

We proceed with approaches using point correspondences and known rotation angles. Du and Brady (1993) as well as Basu (1995) use the optical flow arising from small rotations. Basu applies two independent rotations (pan and tilt), uses contours instead of points, and makes an initial hypothesis on the location of the image center. Du and Brady use the conic sections from the trajectories of points to obtain a more accurate image center estimate. Stein (1995) and Vieville (1994) solve the nonlinear equations of motion arising from a rotation around a fixed axis using known rotation angles. Vieville gives also a solution for the unknown but constant rotation axis.

The most closely related work to ours is the approach by Beardsley et al. (1992). They use the trajectories of the vanishing points arising from the fixed axis rotation of parallel lines on a turntable. Since they can produce a full cycle rotation they obtain a full ellipse on the image. As they do not know the rotation axis one trajectory is not enough to recover the intrinsic parameters. From three ellipses the aspect ratio as well as the image center - intersection of the major axes - can be obtained. The focal length is then computed from the locus of vertices of circular cones that could give rise to the given ellipse. Our approach avoids the use of a special apparatus (turntable) by producing the trajectories with camera rotations. We gain in the sense that the fixed axis is then known
but we pay in accuracy since the arising hyperbolic arcs are of limited extend due to the small amount of feasible rotation.

## 3 From parallel lines to conic sections

To describe the camera model we introduce three coordinate systems: The standard coordinate system $\left(x_{s}, y_{s}, z_{s}\right)$ with origin at the projection center (or optical center or nodal point) and $z_{s}$-axis parallel to the optical axis of the camera, the normalized camera coordinate system $\left(x_{n}, y_{n}\right)$ on the plane $z_{s}=1$ with origin at $(0,0,1)$ of the standard coordinate system, and the image coordinate system ( $x_{p}, y_{p}$ ) where the image data are measured in pixel units. The perspective projection from a point $\left(x_{s}, y_{s}, z_{s}\right)$ in space to a point $\left(x_{n}, y_{n}\right)$ in the plane $z_{s}=1$ is given by $x_{n}=x_{s} / z_{s}$ and $y_{n}=y_{s} / z_{s}$. Thus, a point $\left(x_{n}, y_{n}\right)$ in the normalized system defines the ray $\lambda\left(x_{n}, y_{n}, 1\right)$ through the point in space. The mapping from the normalized to the image coordinate system is an affine transformation

$$
\underbrace{\left(\begin{array}{c}
x_{p}  \tag{1}\\
y_{p} \\
1
\end{array}\right)}_{\boldsymbol{x}_{p}}=\underbrace{\left(\begin{array}{ccc}
s_{x} & s_{x y} & x_{0} \\
0 & s_{y} & y_{0} \\
0 & 0 & 1
\end{array}\right)}_{S} \underbrace{\left(\begin{array}{c}
x_{n} \\
y_{n} \\
1
\end{array}\right)}_{\boldsymbol{x}_{n}}
$$

The task of intrinsic calibration is to correspond ray directions to pixel positions. Since a ray direction is given by a point in the normalized coordinate system intrinsic calibration means recovering the affine transformation defined above. The parameter pair $\left(x_{0}, y_{0}\right)$ is the image center (or principal point) defined as the intersection of the optical axis with the image plane in image coordinates (pixel units). The parameter $s_{y}$ is equal to the focal length divided by the vertical dimension of a cell in the CCD chip. The parameter $s_{x}$ is equal to the focal length divided by the horizontal dimension of a cell and multiplied by the sampling ratio between the digitizer and the CCD-signal. The parameter $s_{x y}$ counts for non-orthogonality of the pixel grid and will be regarded as negligible (see (Faugeras, 1995) for its interpretation). Furthermore, we do not consider in this study non-linear radial distortions that appear in case of wide-angle lenses.

The description given above is not sufficient to model projections of points or lines at infinity, therefore we introduce homogeneous coordinates ( $\tilde{x}_{s}, \tilde{y}_{s}, \tilde{z}_{s}, \tilde{w}_{s}$ ) in $\mathbb{P}^{3}$ for the standard coordinate system, $\tilde{\boldsymbol{x}}_{n}=\left(\tilde{x}_{n}, \tilde{y}_{n}, \tilde{w}_{n}\right)$ in $\mathbb{P}^{2}$ for the normalized, and $\tilde{\boldsymbol{x}}_{p}=\left(\tilde{x}_{p}, \tilde{y}_{p}, \tilde{w}_{p}\right)$ also in $\mathbb{P}^{2}$ for the image coordinate system. We obtain always the inhomogeneous coordinates by dividing by the last homogeneous coordinate if the latter is not zero. The equations $\tilde{w}_{s}=0$ and $\tilde{w}_{n}=0$ define the plane at infinity in $\mathbb{P}^{3}$ and the line at infinity in $\mathbb{P}^{2}$, respectively. As a 2 D affine transformation leaves the line at infinity invariant its mapping on the image coordinate system is also $\tilde{w}_{p}=0$. We introduce now the projection equation for homogeneous coordinates:

$$
\underbrace{\left(\begin{array}{c}
\tilde{x}_{p} \\
\tilde{y}_{p} \\
\tilde{w}_{p}
\end{array}\right)}_{\tilde{\boldsymbol{x}}_{p}}=\underbrace{\left(\begin{array}{ccc}
s_{x} & s_{x y} & x_{0} \\
0 & s_{y} & y_{0} \\
0 & 0 & 1
\end{array}\right)}_{S} \underbrace{\left(\begin{array}{c}
\tilde{x}_{n} \\
\tilde{y}_{n} \\
\tilde{w}_{n}
\end{array}\right)}_{\tilde{\boldsymbol{x}}_{n}} \text { and } \underbrace{\left(\begin{array}{c}
\tilde{x}_{n} \\
\tilde{y}_{n} \\
\tilde{w}_{n}
\end{array}\right)}_{\tilde{\boldsymbol{x}}_{n}}=\left(\begin{array}{cccc}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0
\end{array}\right) \underbrace{\left(\begin{array}{c}
\tilde{x}_{s} \\
\tilde{y}_{s} \\
\tilde{z}_{s} \\
\tilde{w}_{s}
\end{array}\right)}_{\tilde{\boldsymbol{x}}_{s}} .
$$

A set of parallel lines with direction $\left(l_{x}, l_{y}, l_{z}\right)$ in projective space meet at the point of the plane at infinity $\left(l_{x}, l_{y}, l_{z}, 0\right)$. The projection of this point on the image plane is the vanishing point - the intersection of the projections of the parallel lines. Its normalized homogeneous coordinates are $\left(l_{x}, l_{y}, l_{z}\right)$ (easily derived from (2)) and its position on the image plane is at infinity if $l_{z}=0$.

Suppose now that a camera moves from pose 1 to pose 2 with translation vector $\boldsymbol{T}$ and a rotation matrix $R$ :

$$
\left(\begin{array}{c}
\tilde{x}_{s 1}  \tag{2}\\
\tilde{y}_{s 1} \\
\tilde{z}_{s 1} \\
\tilde{w}_{s 1}
\end{array}\right)=\left(\begin{array}{cccc} 
& R & & \boldsymbol{T} \\
0 & 0 & 0 & 1
\end{array}\right)\left(\begin{array}{c}
\tilde{x}_{s 2} \\
\tilde{y}_{s 2} \\
\tilde{z}_{s 2} \\
\tilde{w}_{s 2}
\end{array}\right)
$$

We can see that if a point lies at infinity then only the rotation affects the new position of the point. By applying (2) and $\tilde{w}_{s 1}=\tilde{w}_{s 2}=0$ we obtain the transformation of the point in normalized

$$
\begin{equation*}
\tilde{\boldsymbol{x}}_{n 1}=R \tilde{\boldsymbol{x}}_{n 2} \tag{3}
\end{equation*}
$$

as well as image coordinates

$$
\begin{equation*}
\tilde{\boldsymbol{x}}_{p_{1}}=S R S^{-1} \tilde{\boldsymbol{x}}_{p_{2}} \tag{4}
\end{equation*}
$$

Both transformations are invertible, hence they are collineations in $\mathbb{P}^{2}$. They express the transformation of an image point in case of pure rotation and finite 3D-point depth or in case of arbitrary motion and a 3D-point at infinity. The second equation in (4) gives two constraints in inhomogeneous coordinates used by all approaches (see for example (Hartley, 1994; Vieville, 1994)) for estimation of intrinsic parameters from rotations.

Let us suppose that the camera rotates around a fixed axis $\boldsymbol{N}(\|\boldsymbol{N}\|=1)$. Then the trajectory of the 3D-point in the standard coordinate system is a circle. The resulted cone in the standard coordinate system with vertex at the optical center (Fig. 1) and axis $\boldsymbol{N}$ reads

$$
\begin{equation*}
\left(\boldsymbol{N}^{T} \boldsymbol{x}_{s}\right)^{2}=g^{2}\left\|\boldsymbol{x}_{s}\right\|^{2} \tag{5}
\end{equation*}
$$

with $\boldsymbol{x}_{s}^{T}=\left(x_{s}, y_{s}, z_{s}\right)$ and $g$ equal to the cosine of the opening angle of the cone. Since $\boldsymbol{x}_{n}=\left(1 / z_{s}\right) \boldsymbol{x}_{s}$ the equation in normalized coordinates reads

$$
\begin{equation*}
\left(\boldsymbol{N}^{T} \boldsymbol{x}_{n}\right)^{2}=g^{2}\left\|\boldsymbol{x}_{n}\right\|^{2} \tag{6}
\end{equation*}
$$

and after rearranging terms and transforming to image coordinates

$$
\begin{equation*}
\boldsymbol{x}_{p}^{T} S^{-1 T}\left(\boldsymbol{N} \boldsymbol{N}^{T}-g^{2} I\right) S^{-1} \boldsymbol{x}_{p}=0 \tag{7}
\end{equation*}
$$

which is as expected the equation of a conic section in $\mathbb{R}^{2}$.
There are two degenerate cases for this conic section. We will describe them for our application where the conic sections are trajectories of vanishing points. The first appears when the parallel lines giving rise to the vanishing point are perpendicular to the rotation axis. The cone degenerates to a plane and the trajectory of the vanishing point is a line in the image plane. The second degenerate case is when the direction of the parallel lines
is parallel to the rotation axis. The cone degenerates to a line and the conic section to a point at infinity.

The five coefficients of the conic section depend on the two parameters of the rotation axis, the four intrinsic parameters, and the parameter $g$ which depends on which point is tracked in the scene. As already shown (Beardsley et al., 1992) more than one conic section arising from a different rotation axis is needed to recover the intrinsic parameters. Instead, we will study here the special cases where the rotation axis is known. Using one camera of the binocular head in Fig. 2 we perform one rotation about the tilt axis $x$ and one rotation about the vergence axis $y$ - we will call pan axis $y$.


Figure 2. The camera is rotated with angle $\phi$ about the tilt axis $x$ and with angle $\theta$ about the vergence axis $y$.

For the rotation about the $x$-axis we set $\boldsymbol{N}^{T}=(1,0,0)$ in (7) and obtain

$$
\begin{equation*}
\frac{1-g^{2}}{g^{2}} \frac{\left(x_{p}-x_{0}\right)^{2}}{s_{x}^{2}}-\frac{\left(y_{p}-y_{0}\right)^{2}}{s_{y}^{2}}=1 \tag{8}
\end{equation*}
$$

The center of the hyperbola coincides with the center of the image and the length of the semi-axis in $y$-direction is equal the scale factor $s_{y}$ - see also Fig. 8f. The scaling factor $s_{x}$ is coupled with the cosine of the cone's opening angle and cannot recovered from (8). The ordinate $y_{0}$ of the center gives the symmetry axis of the hyperbola whereas $x_{0}$ and $s_{y}$ fix the position and the slope of the asymptotes, respectively. It will turn out in the experiments that the abscissa $x_{0}$ of the center of the hyperbola cannot be reliably recovered since the position of the asymptotes can not be sufficiently constrained.

For the rotation about the $y$-axis we set $\boldsymbol{N}^{T}=(0,1,0)$ in (7) and obtain

$$
\begin{equation*}
-\frac{\left(x_{p}-x_{0}\right)^{2}}{s_{x}^{2}}+\frac{\left(y_{p}-y_{0}\right)^{2}}{s_{y}^{2}} \frac{1-g^{2}}{g^{2}}=1 . \tag{9}
\end{equation*}
$$

Similar to the $x$-axis rotation $\left(x_{0}, y_{0}\right)$ is the center of the hyperbola - see also Fig. 8c- and only $x_{0}$ and $s_{x}$ can be estimated reliably.

The algorithm is broken into two steps:

1. Rotate around the $x$-axis and fit a conic to the data points. From the coefficients of the conic and (8) obtain $y_{0}$ and $s_{y}$.
2. Rotate around the $y$-axis and fit a conic to the data points. From the coefficients of the conic compute center and the axes of the hyperbola to recover $x_{0}$ and $s_{x}(9)$.

## 4 Statistical estimation

The computation of the intersection of lines in the image and in particular the fitting of a conic need special care regarding error treatment (see also (Kanatani, 1993)).

Let $x_{p} \cos \eta_{i}+y_{p} \sin \eta_{i}-d_{i}=0$ be the equation of the $i$-th line in the image. The line parameters $(\eta, d)$ are obtained with the Hough-Transform. To compute the line intersection point we apply Maximum-Likelihood estimation which maximizes the density

$$
e^{-\sum_{i=1}^{N} \frac{\epsilon_{i}^{2}}{2 \sigma \epsilon_{i}^{2}}} \quad \text { where } \quad \epsilon_{i}=x_{p} \cos \eta_{i}+y_{p} \sin \eta_{i}-d_{i}
$$

and $\sigma_{\epsilon}^{2}$ is the error variance approximately equal to the variance $\sigma_{d}^{2}$ of the distance from the origin to the line according to the applied resolution of the Hough space. The MLE yields a least squares minimization of the linear system

$$
\left(\begin{array}{cc}
\cos \eta_{i} & \sin \eta_{i}
\end{array}\right)\binom{x_{p}}{y_{p}}=d_{i} \quad i=1 \ldots N .
$$

The covariance matrix of the estimate reads

$$
\Sigma=\left(\begin{array}{cc}
\sigma_{x}^{2} & \sigma_{x y}^{2}  \tag{10}\\
\sigma_{x y}^{2} & \sigma_{y}^{2}
\end{array}\right)=\sigma_{d}^{2}\left(\begin{array}{cc}
\sum_{i=1}^{N} \cos ^{2} \eta_{i} & \sum_{i=1}^{N} \cos \eta_{i} \sin \eta_{i} \\
\sum_{i=1}^{N} \cos \eta_{i} \sin \eta_{i} & \sum_{i=1}^{N} \sin ^{2} \eta_{i}
\end{array}\right)^{-1}
$$

The largest eigenvalue of $\Sigma$ reads

$$
\lambda_{\max }(\Sigma)=\frac{2 \sigma_{d}^{2}}{\left(N-\sqrt{N^{2}-4 \sum_{i=1}^{N} \sum_{j=i+1}^{N} \sin ^{2}\left(\eta_{i}-\eta_{j}\right)}\right)} .
$$

We observe that the largest eigenvalue increases with decreasing deviation between the line directions and becomes infinite when all lines are parallel. Line directions tend to be parallel when the effective field of view is small and the intersection point is far away from the center. The eigenvector of the covariance matrix corresponding to the largest eigenvalue gives the direction of maximal uncertainty. It can be shown for the case of two lines that this direction is parallel to the angle bisector of the lines which can be intuitively generalized to the average direction in case of $N$ lines. This covariance matrix is propagated to the next step of the conic fitting.

The equation of the conic (7) can be written as a quadratic form $x_{p}^{T} A x_{p}=0$ where A is a symmetric matrix with $A_{12}=A_{21}=0$ because the rotation axis $\boldsymbol{N}$ in (7) is either $(1,0,0)$ or $(0,1,0)$. Decoupling the vanishing point coordinates from the unknowns we obtain $\boldsymbol{d}^{T} \boldsymbol{a}=0$ with

$$
\boldsymbol{d}^{T}=\left(\begin{array}{lllll}
x_{p}^{2} & y_{p}^{2} & x_{p} & y_{p} & 1 \tag{11}
\end{array}\right)
$$

and

$$
\boldsymbol{a}^{T}=\left(\begin{array}{lllll}
A_{11} & A_{22} & 2 A_{13} & 2 A_{23} & A_{33} \tag{12}
\end{array}\right)
$$

which leads to the weighted least squares minimization of

$$
\begin{equation*}
\sum_{i=1}^{M} \frac{\left(\boldsymbol{d}_{i}^{T} \boldsymbol{a}\right)^{2}}{\operatorname{Var}\left[\boldsymbol{d}_{i}^{T} \boldsymbol{a}\right]} \tag{13}
\end{equation*}
$$

If we exactly compute the variance $\operatorname{Var}\left[\boldsymbol{d}^{T} \boldsymbol{a}\right]$ we will obtain a nonlinear minimization problem because the unknown $\boldsymbol{a}$ will appear in the denominator of every summation term. Therefore, we apply approximate weights equal to the trace $\operatorname{tr}\left(\Sigma_{i}\right)$ of the covariance matrix of the $i$-th vanishing point.

As the vector of unknowns $\boldsymbol{a}$ can be computed up to a scaling factor we have to introduce a constraint if we do not use the exact weightings in (13) where the coefficients appear in the denominators. As Kanatani (1993) argues minimizing with respect to translation and rotation invariance - constraint $A_{11}^{2}+A_{22}^{2}=1$ - does not apply in the case of points with anisotropic noise and unequal weightings. He proposes the usual constraint $\|\boldsymbol{a}\|=1$ leading to the minimization

$$
\begin{equation*}
\min _{\|\boldsymbol{a}\|=1} \boldsymbol{a}^{T} C \boldsymbol{a} \quad \text { with } \quad C=\sum_{i=1}^{M} \frac{1}{\operatorname{tr}\left(\Sigma_{i}\right)} \boldsymbol{d}_{i} \boldsymbol{d}_{i}^{T} . \tag{14}
\end{equation*}
$$

The solution is the eigenvector of $C$ corresponding to the smallest eigenvalue.
The matrix $C$ is corrupted by noise $\delta C$ since it contains the noisy measurements of the vanishing points $\left(x_{p}, y_{p}\right)$. Assuming that the noise ( $\left.\delta x_{p}, \delta y_{p}\right)$ in the points is anisotropic Gaussian with zero mean and covariance matrix $\Sigma$ given in (10) the expectation of $C+\delta C$ reads

$$
E[C+\delta C]=\sum_{i=1}^{M} E\left[C_{i}+\delta C_{i}\right]
$$

To compute $E\left[C_{i}+\delta C_{i}\right]$ for one point we temporarily omit the index $i$ in the point coordinates as well as the point variances

$$
E\left[C_{i}+\delta C_{i}\right]=\frac{1}{\operatorname{tr}(\Sigma)} E\left[\left(\begin{array}{ccc}
\left(x_{p}+\delta x_{p}\right)^{4} & \left(x_{p}+\delta x_{p}\right)^{2}\left(y_{p}+\delta y_{p}\right)^{2} & \cdots \\
\left(x_{p}+\delta x_{p}\right)^{2}\left(y_{p}+\delta y_{p}\right)^{2} & \left(y_{p}+\delta y_{p}\right)^{4} & \cdots \\
\vdots & \vdots &
\end{array}\right)\right]
$$

After computation of the joint moments up to fourth order of $\delta x_{p}$ and $\delta y_{p}$ the expectation $E\left[\delta C_{i}\right]$ reads

$$
E\left[\delta C_{i}\right]=\frac{1}{\operatorname{tr}(\Sigma)}\left(\begin{array}{cccc}
C_{2 x 2} & & C_{2 x 3} &  \tag{15}\\
& \sigma_{x}^{2} & \sigma_{x y}^{2} & 0 \\
C_{2 x 3}^{T} & \sigma_{x y}^{2} & \sigma_{y}^{2} & 0 \\
& 0 & 0 & 0
\end{array}\right)
$$

where

$$
C_{2 x 2}=\left(\begin{array}{cc}
6 x_{p}^{2} \sigma_{x}^{2}+3 \sigma_{x}^{4} & x_{p}^{2} \sigma_{y}^{2}+y_{p}^{2} \sigma_{x}^{2}+4 x_{p} y_{p} \sigma_{x y}^{2}+\sigma_{x}^{2} \sigma_{y}^{2}+2 \sigma_{x y}^{4} \\
x_{p}^{2} \sigma_{y}^{2}+y_{p}^{2} \sigma_{x}^{2}+4 x_{p} y_{p} \sigma_{x y}^{2}+\sigma_{x}^{2} \sigma_{y}^{2}+2 \sigma_{x y}^{4} & 6 y_{p}^{2} \sigma_{y}^{2}+3 \sigma_{y}^{4}
\end{array}\right)
$$

and

$$
C_{2 x 3}=\left(\begin{array}{ccc}
3 x_{p} \sigma_{x}^{2} & 2 x_{p} s_{x y}^{2}+y_{p} \sigma_{x}^{2} & \sigma_{x}^{2} \\
2 y_{p} \sigma_{x y}^{2}+x_{p} \sigma_{y}^{2} & 3 y_{p} \sigma_{y}^{2} & \sigma_{y}^{2}
\end{array}\right)
$$

From the perturbation theorem for eigenvectors (Kanatani, 1993; Golub \& van Loan, 1989) of symmetric matrices we can derive the bias of the solution $\hat{\boldsymbol{a}}$ as

$$
\begin{equation*}
E[\delta \hat{\boldsymbol{a}}]=-\left(\sum_{j=1}^{4} \frac{\boldsymbol{u}_{j} \boldsymbol{u}_{j}^{T}}{\lambda_{j}}\right) E[\delta C] \hat{\boldsymbol{a}} \tag{16}
\end{equation*}
$$

where $\left(\lambda_{j}, \boldsymbol{u}_{j}\right)$ are the four remaining eigenvalue-eigenvector pairs. We observe that the bias is equal zero only if the expectation of $\delta C$ vanishes. Unfortunately, the expectation $E[\delta C]$ is the sum of the matrices $E\left[\delta C_{i}\right]$ equal for every point to the second term of (15) and does not vanish. Hence, to eliminate the bias we have to subtract the expectation $E\left[\delta C_{i}\right]$ in (15) from every data matrix $C_{i}$.

The final minimization that yields an unbiased estimate reads

$$
\begin{equation*}
\min _{\|\boldsymbol{a}\|=1} \boldsymbol{a}^{T}\left(\sum_{i=1}^{M} \frac{1}{\operatorname{tr}\left(\Sigma_{i}\right)} \boldsymbol{d}_{i} \boldsymbol{d}_{i}^{T}-E\left[\delta C_{i}\right]\right) \boldsymbol{a} \tag{17}
\end{equation*}
$$

As simulation tests by (Fitzgibbon \& Fisher, 1995) showed the Kanatani procedure is superior for data from small portions of the conic. The same authors experimentally showed that this algorithm does not always detect the desired conic kind - here hyperbola - for high noise levels. However, such estimates were observed in our experiments in a much smaller extent than the one given in (Fitzgibbon \& Fisher, 1995). In this section we described how the covariance matrix of the estimated vanishing points is introduced in the conic fitting step in order to properly weigh the data as well as to eliminate the bias in the estimates of the conic coefficients.

## 5 Ep eriments

We begin with experimental results on synthetic data. The image lines used are noise corrupted images of the lines on a cube covered with a checker-board similar pattern. Noise is added to the image points before line fitting. Computation of the vanishing points and the conic coefficients is carried out as described in the previous section. We present the results only in the computation of $x_{0}$ and $s_{x}$ from a rotation around the $y$-axis. The computation of $y_{0}$ and $s_{y}$ using synthetic data is exactly symmetric.

We compare our approach with the two-step approach in (Li, 1994). In this approach the image center $\left(x_{0}, y_{0}\right)$ is computed from the intersection of the radial point trajectories arising from a varying focal length (zooming). Then, the scaling factors ( $s_{x}, s_{y}$ ) are estimated from three vanishing points arising from three mutually orthogonal sets of parallel lines. Of course we expect that this approach exhibits a much better performance than ours since three sets of parallel lines and over twenty zoom point trajectories are exploited.

We first inspect the sensitivity to the noise level in the image line parameters. The rotation amount is 120 degrees around the $y$-axis. We observe (Fig. 3) that the relative error in the image center (x-coordinate $x_{0}$ ) is an order of magnitude lower than the error in the scaling factor $s_{x}$.

To better explain the sensitivity behavior we proceed by varying the most important factor of our approach: the total amount of rotation. We see in Fig. 4(a-b) the distribution of the vanishing points for rotation amounts of 120 and 20 degrees. As expected the relative error in the intrinsics (Fig $4(c, d)$ ) decreases with growing rotation amount. The reason is that when the region expanded by the same number of points grows the hyperbola to be fitted is more constrained. Again we observe an order of magnitude difference between error in the image center and the scaling factor.

We next inspect how the position of the vanishing points on the hyperbola affects the computation. This factor called symmetry index is determined by the median angle during the rotational course. Symmetry index 0 means rotating from -40 to 40 degrees


Figure 3. The relative error in $x_{0}(a)$ and in $s_{x}$ (b) as a function of the standard deviation of noise in the line parameters. The amount of rotation was 120 degrees. The lower curve shows the error of the zoom-vanishing point algorithm.
whereas symmetry index 45 means moving from 5 to 85 degrees. We see the two extremal distributions in Fig. 5 (a) and (b). The total amount of rotation in this experiment is 80 degrees. The relative error in the x-coordinate of the image center increases if symmetry is violated. Since $x_{0}$ gives the position of the transverse symmetry axis of the hyperbola it is expected that asymmetric distribution destabilizes the transverse axis. On the other hand, the error in the scaling factor $s_{x}$ decreases (Fig. 5 (d)) as more points lie laterally on the hyperbola because the slope of the asymptote is constrained.

In the fourth experiment we tested the behavior of the algorithm between the two degenerate cases. The relative errors are illustrated in Fig. 6 as functions of the angle between rotation axes and parallel line directions. At the two extremes the hyperbola is a line (angle $=0$ ) and a point in infinity (angle $=90$ ), respectively. The error decreases while the degenerate line becomes a conic and then it grows with increasing angle after 45 degrees because the vanishing points go even further away from the image.

In the fifth experiment we tested the behavior under weak perspective. The relative errors in Fig. 7 are plotted as functions of the distance to a given set of parallel lines on an object. When the object moves away the vanishing point remains in the same position. However, its computation becomes unstable because the images of the parallel lines span a very small visual angle so that the uncertainty in the intersection point increases.

In all the experiments our algorithm could never compete with the zoom-vanishing points approach (Li, 1994) due to the much less amount of information used here. We tested both steps of our algorithm with two sequences of the image of a wall obtained with one of the cameras of an active binocular camera mount. The left column in Fig. 8 shows a pan movement and the right column a tilt movement. The image lines were detected using the Hough-transform applied on the edge image. In Fig. 8 (c) and (f) we see the trajectories of the vanishing points arising from a 120 degrees rotation. Ground truth for the camera is not known but the approach ( $\mathrm{Li}, 1994$ ) applied before the experiment using a calibration cube serves as a reference. The relative error with respect to this reference was $5.0 \%$ for $x_{0}, 9.8 \%$ for $y_{0}, 2.7 \%$ for $s_{x}$, and $5.3 \%$ for $s_{y}$. With respect to the data used and the fact that the rotation was not 360 degrees like in other approaches the results are more than satisfactory.


Figure 4. Vanishing points and fitted hyperbolae for a rotation amount of 120 degrees (a) and 20 degrees (b), respectively. The rectangle is the image boarder drawn to illustrate the relative position of the vanishing points. The relative error in $x_{0}$ (c) and $s_{x}$ (d) as a function of the rotation amount. The dashed line is the error in the zoom-vanishing point approach which does not use rotations.

## 6 Conclusion

We presented an approach for estimating the intrinsic parameters of a camera able to rotate around the tilt and pan axes. The intrinsic parameters can be obtained from the coefficients of two conic sections. No calibration pattern was used. The use of vanishing points instead of image point matches enables calibration during translation of the observer. Detection of parallel lines is more stable then image feature detection and tracking although the recognition of a group of parallels in a cluttered scene necessitates a priori knowledge on the scene.

The intuition and the experimental tests show that the most critical factors of our calibration are the amount of rotation and the symmetry of the distribution of the points along the hyperbola. Synthetic experiments show that the relative position of the optical axis, the rotation axis, and the direction of parallel lines affects the estimation accuracy. A method should be elaborated that a priory finds the optimal pose of the camera system


Figure 5. The vanishing points and the fitting conics for symmetry index 0 degrees (a) and 45 degrees (b). The relative error in $x_{0}$ (c) and $s_{x}$ (d) as a function of the symmetry index. As the symmetry index varies from 0 to 45 degrees the set of vanishing points is shifted to the left of the hyperbola. For comparison is shown the constant error in the zoom-vanishing point approach (dashed line).
during rotation.

## Acknowledgements

We are grateful to the anonymous reviewers who contributed with their constructive comments to a substancial improvement of the soundness and readability of this paper.

## References

Basu, A. (1995). Active calibration of cameras. IEEE Trans. Systems, Man, and Cybernetics, 25, 256-265.

Beardsley, P., Murray, D., \& Zisserman, A. (1992). Camera calibration using multiple images. In Proc. Second European Conference on Computer Vision, pp. 312-320. Santa Margerita, Italy, May 23-26, G. Sandini (Ed.), Lecture Notes in Computer Science 588, Springer-Verlag, Berlin et al.


Figure 6. The relative error in $x_{0}$ (a) and $s_{x}$ (b) as a function of the angle between rotation axis and parallel line direction. When both directions are parallel (zero angle) the hyperbola degenerates to a straight line through the origin. When the rotation axis is vertical to the parallels (angle $=90 \mathrm{deg}$ ) the projections of the parallels do not intersect at all.


Figure 7. The relative error in $x_{0}(a)$ and $s_{x}(b)$ as a function of the distance between camera and the lines in 3D. The more distant the lines are the weaker is the perspective. The relative error for the scaling factor increases also when using the zoom-vanishing point algorithm which makes use of the images of three sets of parallel lines.

Caprile, B., \& Torre, V. (1990). Using vanishing points for camera calibration. International Journal of Computer Vision, 4, 127-140.

Dron, L. (1993). Dynamic camera self-calibration from controlled motion sequences. In IEEE Conf. Computer Vision and Pattern Recognition, pp. 501-506. New York, NY, June 15-17.

Du, F., \& Brady, M. (1993). Self-calibration of the intrinsic parameters of cameras for active vision systems. In IEEE Conf. Computer Vision and Pattern Recognition, pp. 477-482. New York, NY, June 15-17.

Echigo, T. (1990). A camera calibration technique using three sets of parallel lines. Machine Vision and Applications, 3, 159-167.

Faugeras, O. (1992). What can be seen in three dimensions with an uncalibrated stereo rig. In Proc. Second European Conference on Computer Vision, pp. 563-578. Santa Margerita, Italy, May 23-26, G. Sandini (Ed.), Lecture Notes in Computer Science 588, Springer-Verlag, Berlin et al.

Faugeras, O. (1995). Stratification of three-dimensional vision: projective, affine, and metric representations. Journal Opt. Soc. Am. A, 12, 465-484.

Fitzgibbon, A., \& Fisher, R. (1995). A buyer's guide to conic fitting. In Sixth British Machine Vision Conference, Birmingham, UK, Sept. 11-14.

Golub, G., \& van Loan, C. (1989). Matrix Computations. Second Edition, The Johns Hopkins University Press, Baltimore, Maryland.

Hartley, R. (1994). Self-calibration from multiple views with a rotating camera. In Proc. Third European Conference on Computer Vision, pp. 471-478. Stockholm, Sweden, May 2-6, J.O. Eklundh (Ed.), Springer LNCS 800.

Kanatani, K. (1993). Geometric Computation for Machine Vision. Oxford University Press, Oxford, UK.

Li, M. (1994). Camera calibration of a head-eye system for active vision. In Proc. Third European Conference on Computer Vision, pp. 543-554 Stockholm, Sweden, May 2-6, J.O. Eklundh (Ed.), Springer LNCS 800.

Luong, Q., \& Vieville, T. (1994). Canonic representations for the geometries of multiple projective views. In Proc. Third European Conference on Computer Vision, pp. 589-599. Stockholm, Sweden, May 2-6, J.O. Eklundh (Ed.), Springer LNCS 800.

Lutton, E., Maitre, H., \& Lopez-Krahe, J. (1994). Contribution to the determination of vanishing points using Hough transform. IEEE Trans. Pattern Analysis and Machine Intelligence, 16, 430-438.

Maybank, S., \& Faugeras, O. (1992). A theory of self-calibration of a moving camera. International Journal of Computer Vision, 8, 123-151.

Palmer, P., Petrou, M., \& Kittler, J. (1993). An optimisation approach to improving the accuracy of the Hough transform: Plane orientations from skew symmetry. In Proc. Int. Conf. on Computer Vision, pp. 529-533 Berlin, Germany, May 11-14.

Stein, G. (1995). Accurate internal camera calibration using rotation, with analysis of sources of errors. In Proc. Int. Conf. on Computer Vision, pp. 230-236. Boston, MA, June 20-23.

Tsai, R. (1986). An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. In IEEE Conf. Computer Vision and Pattern Recognition, pp. 364-374. Miami Beach, FL, June 22-26.

Vieville, T. (1994). Auto-calibration of visual sensor parameters on a robotic head. Image and Vision Computing, 12, 227-237.


Figure 8. The left column describes a pan-rotation around the $y$-axis from left to right. Images (a) and (b) are the first and the last in the sequence, respectively. The regarded set of parallel lines is the set with vanishing point on the upper right of image (a). Fig. (c) shows the vanishing points moving from right to left and the fitted hyperbolae. In the right column we show the tilt-rotation around the x-axis from bottom to top. Images (d) and (e) are the first and the last in the sequence, respectively. The regarded set of parallel lines is the set with vanishing point on the left of image (d). The vanishing point in Fig. (f) is moving from top to bottom.

