ISPRS Journal of Photogrammetry and Remote Sensing 74 (2012) 202–217

Contents lists available at SciVerse ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Estimating the fundamental matrix under pure translation and radial distortion

Carsten Steger *

MVTec Software GmbH, Neherstraße 1, 81675 München, Germany

ARTICLE INFO

Article history:
Received 20 October 2011
Received in revised form 8 September 2012
Accepted 10 September 2012

Keywords:
Uncalibrated stereo
Fundamental matrix
Radial distortion
Division model
Minimal solver
Overdetermined solver

ABSTRACT

Four algorithms for estimating the fundamental matrix and radial distortion for the case that the images are related by a pure camera translation are proposed. This restricted class of motions is important in robotics and industrial applications. It is assumed that the radial distortions can be modeled by the division model. First, a linear 3-point minimal solver is derived, which is essential to drive a RANSAC algorithm. The minimal solver is then extended to an overdetermined linear solver, which is essential to compute starting values for the third algorithm, a Gold Standard solver that returns the maximum likelihood estimate fundamental matrix and radial distortion. The three algorithms are combined into a RANSAC algorithm that determines the fundamental matrix and radial distortion automatically. An experimental evaluation determines the characteristics of the algorithms and shows the high accuracy, precision, and robustness that can be achieved. Furthermore, the evaluation shows that using algorithms that determine the fundamental matrix and radial distortion for general camera motions are not a viable alternative for the case of pure camera translations because they are significantly slower and less robust than the algorithms proposed here.

1. Introduction

If two uncalibrated images are related by a pure camera translation, it is well known that the scene can be reconstructed up to an affine transformation, instead of a general projective transformation if the camera has undergone a general motion (Moons et al., 1993; Armstrong et al., 1994; Hartley and Zisserman, 2003, Section 10.4). As is customary (Hartley and Zisserman, 2003, Section 9.3), the term "pure camera translation" implies that the interior orientation of the camera remains constant. We make the same assumption in this paper.

A very important application area where pure camera translations occur frequently is industrial applications in which images of objects to be inspected are acquired by a stationary camera mounted over a conveyor belt. In industrial applications, high accuracy is often mandatory. Using existing algorithms for the pure camera translation case (Moons et al., 1993; Armstrong et al., 1994; Hartley and Zisserman, 2003, Section 10.4), however, does not enable us to achieve high accuracy since these algorithms do not model the radial distortions that are inherent in any lens. Therefore, in this paper we propose several algorithms to determine the fundamental matrix and radial distortion under pure camera translation. We will assume that the radial distortion can be modeled adequately by the division model (Lenz, 1987, 1988; Lenz and Fritsch, 1990; Lanser et al., 1995; Lanser and Zierl, 1995; Fitzgibbon, 2001).

We will first describe and analyze the division model in detail in Section 2. In particular, we will see that the manner in which the division model is used in this paper implies the assumption that the pixels are approximately square (aspect ratio close to one and skew close to zero) and that the principal point lies close to the image center.

For pure camera translations, the fundamental matrix only has two degrees of freedom: the common epipole of the two images (Hartley and Zisserman, 2003, Section 10.4). Our goal is to derive an algorithm that determines the fundamental matrix $\mathtt{F}$ and the radial distortion coefficient λ automatically. This will be done using an adaptation of a standard random sample consensus (RANSAC) algorithm (Fischler and Bolles, 1981; Hartley and Zisserman, 2003, Section 4.7). Thus, in Section 3 we will propose a minimal solver for this problem. Since the minimal solver must determine three parameters, it requires three point correspondences to determine two possible solutions of $\mathtt{F}$ and λ. The 3-point minimal solver extends the list of minimal solvers that have been proposed recently for various problems: the 5-point relative pose problem (Nistér, 2004; Stewénius, 2005; Stewénius et al., 2006; Li and Hartley, 2006), the 2- or 3-point panoramic stitching problem (Brown et al., 2007), the 8-point general fundamental matrix problem with radial distortion (Kúkelová and Pajdla, 2007; Kúkelová et al., 2008), the 9-point general fundamental matrix problem with different radial distortions (Byröd et al., 2008; Kúkelová et al., 2010), and the

* Tel.: +49 89 4576950; fax: +49 89 45769555.
E-mail address: steger@mvtec.com

0924-2716/$ - see front matter © 2012 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS) Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.isprsjprs.2012.09.012

3-point panoramic stitching problem with radial distortion (Byröd et al., 2009). We also note that non-minimal 9-point solvers have been described for the general fundamental matrix problem with radial distortion (Fitzgibbon, 2001; Li and Hartley, 2005).

In an automatic algorithm, we also must handle the non-minimal cases. Therefore, we propose an overdetermined linear solver in Section 4 and a Gold Standard solver in Section 5. Section 6 discusses the reconstruction ambiguity that can be handled with the proposed algorithms, while Section 7 examines different degeneracies that can affect the algorithms. The three algorithms above are integrated in Section 8 into an algorithm that determines $\mathtt{F}$ and λ automatically. Finally, we will carry out a performance evaluation of the proposed algorithms in Section 9.

2. The division model for radial distortion

The division model for radial distortion has been in common use for over two decades in calibrated machine vision applications (Lenz, 1987, 1988; Lenz and Fritsch, 1990; Lanser et al., 1995; Lanser and Zierl, 1995). Based on the work presented by Fitzgibbon (2001), it has recently received significant attention for problems in uncalibrated machine vision applications (Byröd et al., 2008, 2009; Kúkelová and Pajdla, 2007; Kúkelová et al., 2008, 2010; Li and Hartley, 2005; Strand and Hayman, 2005). In this section, we will present a short overview of the division model and its properties.

The division model describes a transformation that occurs in the image plane. We will denote the undistorted points by $\mathbf{x}_u$ and the distorted points by $\mathbf{x}_d$. According to the division model, the distorted points can be undistorted as follows:

$$\mathbf{x}_u = \frac{\mathbf{x}_d}{1 + \lambda \|\mathbf{x}_d\|^2} \tag{1}$$

One of the major advantages of the division model is that the distortion function, i.e., the inverse of (1), can be computed analytically (Lenz, 1987, 1988; Lenz and Fritsch, 1990):

$$\mathbf{x}_d = \frac{2\mathbf{x}_u}{1 + \sqrt{1 - 4\lambda \|\mathbf{x}_u\|^2}} \tag{2}$$

For the division model to hold exactly, it is assumed in the above equations that the center of the radial distortions coincides with the center of the image. Consequently, we will assume that $\mathbf{x}_u$ and $\mathbf{x}_d$ are specified relative to the image center. In practice, we will only assume that the center of the radial distortions is close enough to the image center so that the division model holds with sufficient accuracy. Note that the principal point of the camera is also commonly assumed to be the center of the radial distortions under the division model (Lenz, 1987, 1988; Lenz and Fritsch, 1990; Lanser et al., 1995; Lanser and Zierl, 1995). This assumption is fulfilled for cameras in which the optical axis is perpendicular to the sensor, which is by far the most common configuration in industrial applications. However, this also implies that the proposed algorithm cannot be used for cameras that are based on the Scheimpflug principle. In practice, we only assume that the optical axis is sufficiently close to perpendicular to the sensor so that (1) and (2) can be applied with sufficient accuracy. Furthermore, note that the above formulas implicitly assume that the camera has square pixels and zero skew if the division model is to hold exactly. In practice, we only assume that the pixels are close enough to being square and having zero skew that (1) and (2) can be applied with sufficient accuracy. This is an unimportant restriction since almost all industrial cameras have square pixels with zero skew to a very high degree of accuracy. If we wanted to model non-square pixels and non-zero skew, we would have to multiply the distorted point coordinates by an appropriate 2D affine transformation, i.e., we would have to model these effects after the distortion has happened.

If $\lambda = 0$, the points are not distorted. For $\lambda > 0$, we obtain a pincushion distortion, while for $\lambda < 0$ we obtain a barrel distortion. It should be noted that the value of λ varies according to the scaling of the input points. If the points $\mathbf{x}_u$ and $\mathbf{x}_d$ are specified in pixels, λ has units (1/pixel^2). On the other hand, if the points $\mathbf{x}_u$ and $\mathbf{x}_d$ are assumed to be normalized to the range $[-1,1] \times [-1,1]$, λ becomes a dimensionless quantity. In any event, if the points are scaled by a common scale factor s, λ will be scaled by $1/s^2$. Thus, it is easy to convert between the two conventions.

In the algorithms below, we will need to restrict the allowed values of λ. By looking at the denominator of (1), we can identify a singularity for $1 + \lambda\|\mathbf{x}_d\|^2 = 0$. This corresponds to points in the image being undistorted to the line at infinity. By looking at the limit of $1 + \lambda\|\mathbf{x}_d\|^2 \downarrow 0$, we can see that there is a circle in the image for which the points will be undistorted to the line at infinity. Furthermore, we can see that the case $1 + \lambda\|\mathbf{x}_d\|^2 < 0$ is not physically meaningful. It corresponds to points being undistorted beyond the line at infinity, which is equivalent to a reflection in the line at infinity. Note that this can only happen for $\lambda < 0$. Therefore, we will require

$$1 + \lambda\|\mathbf{x}_d\|^2 > 0 \tag{3}$$

in the algorithms below. To derive a restriction for λ from (3), we will have to specify the location of the circle in the image that is undistorted to the line at infinity. To do so, we will consider the most extreme case that the pinhole model can theoretically handle, i.e., a configuration in which a complete 180° hemisphere is projected into the image plane. Of course, without distortions, the image plane would have to be infinite in this case. We then consider a distortion by the division model that distorts the infinite image plane into a finite area that can be represented in an image. Since the division model is radially symmetric, it compresses the infinite image plane into a circle. To derive meaningful limits for λ, we require that the distorted image plane completely fits into the sensor image, i.e., that it touches the borders of the image along its smaller dimension (typically the top and bottom borders).[1] Thus, the maximum useful value for $\|\mathbf{x}_d\|$ in (3) is given by $\min(w,h)/2$, where w and h denote the width and height of the image, respectively. Therefore, we require

$$\lambda > -\frac{4}{\min(w,h)^2} \tag{4}$$

To derive an upper bound for λ, we can look at the term $1 - 4\lambda\|\mathbf{x}_u\|^2$ in the denominator of (2). If this term is 0, a point $\mathbf{x}_u$ is distorted exactly by a factor of 2, i.e., $\mathbf{x}_d = 2\mathbf{x}_u$. Obviously, we must require

$$1 - 4\lambda\|\mathbf{x}_u\|^2 \geqslant 0 \tag{5}$$

As above, we must specify an upper limit for $\|\mathbf{x}_u\|$ to derive a limit for λ. In this case, we require that the range of points that can be distorted should completely lie within the image. Thus, the maximum distorted point is given by $\mathbf{x}_d = (w/2, h/2)$, which corresponds to the undistorted point $\mathbf{x}_u = (w/4, h/4)$. Therefore, we require

$$\lambda \leqslant \frac{4}{w^2 + h^2} \tag{6}$$

[1] We note that this is similar to a circular fish-eye lens that projects a complete 180° hemisphere into the image. Note that we are not claiming that a circular fish-eye lens can be modeled adequately by the division model. We also note that there are circular fish-eye lenses with a field of view larger than 180°. Nevertheless to derive meaningful bounds for λ, imagining a circular fish-eye lens with a 180° field of view is a useful analogy.

It should be noted that the width and height in (4) and (6) are specified in pixels. Therefore, if the coordinates are scaled by a common factor, w and h must be scaled by the same factor.

3. The minimal solver

The automatic determination of the fundamental matrix $\mathtt{F}$ and radial distortion coefficient λ in Section 8 requires a minimal solver that generates candidate solutions. The problem can be stated as follows: given a minimal number of point correspondences (three in our case) $\mathbf{x}'_d \leftrightarrow \mathbf{x}_d$, find $\mathtt{F}$ and λ such that

$$\mathbf{x}'^{\top}_u \mathtt{F} \mathbf{x}_u = 0 \tag{7}$$

Following Fitzgibbon (2001), we will use (1) to write $\mathbf{x}_u = \mathbf{x}_d + \lambda \mathbf{z}_d$. In homogeneous coordinates, we have

$$\begin{bmatrix} x_u \\ y_u \\ 1 \end{bmatrix} = \begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix} + \lambda \begin{bmatrix} 0 \\ 0 \\ r_d \end{bmatrix} = \begin{bmatrix} x_d \\ y_d \\ 1 + \lambda(x_d^2 + y_d^2) \end{bmatrix} \tag{8}$$

where $r_d = \|\mathbf{x}_d\| = (x_d^2 + y_d^2)$. Thus, we obtain

$$(\mathbf{x}'_d + \lambda \mathbf{z}'_d)^{\top} \mathtt{F} (\mathbf{x}_d + \lambda \mathbf{z}_d) = \mathbf{x}'^{\top}_d \mathtt{F} \mathbf{x}_d + \lambda (\mathbf{z}'^{\top}_d \mathtt{F} \mathbf{x}_d + \mathbf{x}'^{\top}_d \mathtt{F} \mathbf{z}_d) + \lambda^2 \mathbf{z}'^{\top}_d \mathtt{F} \mathbf{z}_d = 0 \tag{9}$$

We are interested in the pure translation case, for which $\mathtt{F}$ has a simple form given by (Hartley and Zisserman, 2003, Section 9.3):

$$\mathtt{F} = [\mathbf{e}]_{\times} = \begin{bmatrix} 0 & -e_z & e_y \\ e_z & 0 & -e_x \\ -e_y & e_x & 0 \end{bmatrix} \tag{10}$$

where $\mathbf{e}$ is the epipole, which for pure translations is identical for both images.

The structures of $\mathtt{F}$, $\mathbf{z}_d$, and $\mathbf{z}'_d$ lead to the result that the term $\lambda^2 \mathbf{z}'^{\top}_d \mathtt{F} \mathbf{z}_d$ in (9) vanishes. By inserting (10) into (9) and collecting terms, we see that each point correspondence leads to a constraint of the form

$$[y_d - y'_d \quad x'_d - x_d \quad y'_d x_d - x'_d y_d]\mathbf{e} + \lambda [r'_d y_d - r_d y'_d \quad r_d x'_d - r'_d x_d \quad 0]\mathbf{e} = 0 \tag{11}$$

By collecting the constraints of three point correspondences, we must solve the following generalized eigenvalue problem:

$$\mathtt{A}\mathbf{e} = -\lambda \mathtt{B} \mathbf{e} \tag{12}$$

where the design matrices $\mathtt{A}$ and $\mathtt{B}$ are obtained by stacking the corresponding rows of (11):

$$\mathtt{A} = \begin{bmatrix} y_{d,1} - y'_{d,1} & x'_{d,1} - x_{d,1} & y'_{d,1} x_{d,1} - x'_{d,1} y_{d,1} \\ y_{d,2} - y'_{d,2} & x'_{d,2} - x_{d,2} & y'_{d,2} x_{d,2} - x'_{d,2} y_{d,2} \\ y_{d,3} - y'_{d,3} & x'_{d,3} - x_{d,3} & y'_{d,3} x_{d,3} - x'_{d,3} y_{d,3} \end{bmatrix} \tag{13}$$

$$\mathtt{B} = \begin{bmatrix} r'_{d,1} y_{d,1} - r_{d,1} y'_{d,1} & r_{d,1} x'_{d,1} - r'_{d,1} x_{d,1} & 0 \\ r'_{d,2} y_{d,2} - r_{d,2} y'_{d,2} & r_{d,2} x'_{d,2} - r'_{d,2} x_{d,2} & 0 \\ r'_{d,3} y_{d,3} - r_{d,3} y'_{d,3} & r_{d,3} x'_{d,3} - r'_{d,3} x_{d,3} & 0 \end{bmatrix} \tag{14}$$

Since the matrix $\mathtt{B}$ is singular, one of the eigenvalues will be ∞, which will cause no conceptual or numerical difficulties, but causes a performance degradation on current computers. By letting $\mu = \lambda^{-1}$, it is more efficient to solve the equivalent generalized eigenvalue problem

$$-\mathtt{B}\mathbf{e} = \mu \mathtt{A} \mathbf{e} \tag{15}$$

In this problem, one eigenvalue μ will always be 0 and can, therefore, be ignored. The remaining two eigenvalues and the corresponding eigenvectors form the potential solutions for μ and $\mathbf{e}$, respectively, which will be checked for validity as described below.

When the minimal solver constructs the 3×3 design matrices $\mathtt{A}$ and $\mathtt{B}$, an appropriate normalization of the point coordinates is important for numerical reasons. The essential requirement is that all the quantities involved in the problem specification (including λ) have approximately the same size. The discussion in Section 2 shows that this will be the case if the point coordinates are scaled to have an absolute value smaller than one, i.e., by a scale factor of $2/\max(w,h)$. The algorithm next computes the eigenvalues μ and eigenvectors $\mathbf{e}$ of (15). It then checks the nonzero eigenvalues μ. If they are imaginary, the problem has no solution. If they are both real, the two solutions $\lambda_{1,2} = \mu_{1,2}^{-1}$ are checked whether they fall within the allowable range given by (4) and (6). The permissible solutions for λ, along with their corresponding epipoles $\mathbf{e}$, are returned and used in the algorithm of Section 8.

Finally, it should be noted that the minimal solver has some degenerate configurations that will be discussed in Section 7.

4. The overdetermined linear solver

The algorithm in the previous section is the minimal solver. It expects exactly three point correspondences. In real applications, we would like to determine the fundamental matrix from many more point correspondences to achieve robustness to noise and greater precision and accuracy. The minimal solver cannot be used for this purpose since the generalized eigenvalue problem expects square matrices. For speed reasons, we are interested in a linear, i.e., non-iterative, algorithm.

Similar to Fitzgibbon (2001), we can turn the overdetermined problem into a generalized eigenvalue problem by a procedure akin to constructing the normal equations in a linear least-squares problem: we multiply both sides of (15) by $\mathtt{A}^{\top}$ and solve the following problem:

$$-\mathtt{A}^{\top}\mathtt{B}\mathbf{e} = \mu \mathtt{A}^{\top}\mathtt{A}\mathbf{e} \tag{16}$$

Note that $\mathtt{A}^{\top}\mathtt{B}$ is singular because $\mathtt{B}$ is singular. Therefore, like the minimal solver, (16) has at least one zero eigenvalue, which can be ignored.

The remaining two eigenvalues of (16) are potential solutions that need to be further evaluated. If they are imaginary, the problem has no solution. If they are real, we check the corresponding values of λ against the bounds (4) and (6). Typically, only one of the two values will lie within the allowable range, and we construct the corresponding fundamental matrix and return it along with λ.

In rare cases, both values of λ may lie within the allowable range. If this happens, we need to decide which of the two values is correct. Unlike in the minimal solver, in the overdetermined case we have redundancies that we can use to determine which solution is correct. The correct solution can be identified by determining the residual error of the two solutions and returning the solution with the smaller residual error. The obvious solution is to use the residual error definition of Hartley and Zisserman (2003, Section 11.5). To use this definition, we must undistort the points using (1). Then, the residual error can be defined as:

$$\epsilon^2 = \sum_{i=1}^{n} d\left(\mathbf{x}_{u,i}, \mathtt{F}^{\top}\mathbf{x}'_{u,i}\right)^2 + d\left(\mathbf{x}'_{u,i}, \mathtt{F}\mathbf{x}_{u,i}\right)^2 \tag{17}$$

where n is the number of point correspondences. To compute the two distance terms in (17), for each pair of corresponding points we compute their epipolar lines via $\mathbf{l} = \mathtt{F}^{\top}\mathbf{x}'_u$ and $\mathbf{l}' = \mathtt{F}\mathbf{x}_u$, we normalize the lines by dividing them by the Euclidean length of the sub-vector of their first two components, and we compute the distance of the points from the corresponding epipolar line, i.e., $d = \mathbf{l}^{\top}\mathbf{x}_u$ and $d' = \mathbf{l}'^{\top}\mathbf{x}'_u$.

The above approach turns out to be less than optimal, however. The reason is that it may happen that one of the candidate solutions lies close to the boundaries given by (4) and (6), while the other value may have significantly smaller absolute value. Undistorting the points with an extreme value of λ distorts the distance measurements away from the center of the image significantly. For extreme values of λ, the undistorted distance measurements can become significantly compressed towards the borders of the image. These distance values are compared to the solution with the λ of smaller absolute value, where a significantly smaller distortion of the distance values occurs. Thus, it can happen that the solution with the extreme λ is preferred over the correct solution because of the distance distortion. A solution for this problem is to compute the residual error in distorted image coordinates. This forces both solutions to use the same distance metric. We define the residual error in the original image coordinate system as follows:

$$\epsilon^2 = \sum_{i=1}^{n} d\left(\mathbf{x}_{d,i}, \mathbf{y}_{d,i}\right)^2 + d\left(\mathbf{x}'_{d,i}, \mathbf{y}'_{d,i}\right)^2 \tag{18}$$

Here, $\mathbf{y}_d$ and $\mathbf{y}'_d$ denote the distorted points that correspond to the undistorted points on the epipolar lines $\mathbf{l}$ and $\mathbf{l}'$ that have the minimum distance to $\mathbf{x}_u$ and $\mathbf{x}'_u$, respectively. In detail, we compute the undistorted points and corresponding epipolar lines as above. The undistorted points on the epipolar lines that have the smallest distance from the corresponding points are given by $\mathbf{y}_u = \mathbf{x}_u - d\mathbf{l}$ and $\mathbf{y}'_u = \mathbf{x}'_u - d'\mathbf{l}'$, respectively. We then distort the points via (2) and compute their Euclidean distances to obtain the two distance terms in (18).

We will give more details about the reason for this definition of residual error in Section 8, where a similar problem occurs when counting inliers.

We note that Strand and Hayman (2005) show that the division model distorts lines in the image to circles. Thus, undistorted epipolar lines become distorted epipolar circles. Therefore, an alternative to the above algorithm for computing the residual error would be to compute the epipolar circles that correspond to the distorted point correspondences and to determine the minimum distances of the distorted points from the epipolar circles. We have not pursued this idea further, however.

Like the minimal solver, the overdetermined linear solver has some degenerate configurations that will be discussed in Section 7 below.

Finally, as will be seen by the experimental evaluation in Section 9.1, the overdetermined linear solver returns results that can be significantly biased if the true value of λ is very small and the errors in the points $\mathbf{x}_d$ and $\mathbf{x}'_d$ are sufficiently large. This happens because (16) defines an algebraic error that is biased.[2] This bias leads to a potential instability of the algorithm if the true value of λ is very small. Instability in this context means that the algorithm does not find a solution within the bounds (4) and (6) although the true value of λ lies within these bounds. Therefore, if the algorithm finds no feasible solution (either because the two potential solutions are imaginary or because all solutions have values of λ outside the bounds (4) and (6)), it switches to an overdetermined linear solver for the pure translation case without radial distortions.

5. The Gold Standard solver

For applications that require the highest possible accuracy, a Gold Standard solver is required. This is a maximum likelihood estimate for the fundamental matrix and the radial distortion coefficient, which is obtained by a two-view bundle adjustment. This type of optimization can be solved efficiently by the sparse Levenberg–Marquardt algorithm given in Hartley and Zisserman (2003, Algorithm A6.3).

The problem to be solved can be stated as follows: given $n \geqslant 4$ point correspondences $\left\{\mathbf{x}_{d,i}, \mathbf{x}'_{d,i}\right\}$, determine the fundamental matrix $\mathtt{F}$, the radial distortion coefficient λ, and a set of subsidiary point correspondences $\left\{\hat{\mathbf{x}}_{d,i}, \hat{\mathbf{x}}'_{d,i}\right\}$ that minimize the geometric error

$$\epsilon^2 = \sum_{i=1}^{n} d\left(\mathbf{x}_{d,i}, \hat{\mathbf{x}}_{d,i}\right)^2 + d\left(\mathbf{x}'_{d,i}\hat{\mathbf{x}}'_{d,i}\right)^2 \tag{19}$$

subject to $\hat{\mathbf{x}}'^{\top}_{u,i}\mathtt{F}\hat{\mathbf{x}}_{u,i} = 0$, where $\hat{\mathbf{x}}_{d,i}$ and $\hat{\mathbf{x}}'_{d,i}$ are related to $\hat{\mathbf{x}}_{u,i}$ and $\hat{\mathbf{x}}'_{u,i}$ by (2). Hence, we measure the geometric error in the original, distorted image coordinate system and we enforce that the undistorted subsidiary point correspondences satisfy the epipolar constraint. Like in the minimal and overdetermined linear solvers, we normalize the point coordinates by scaling them by a factor of $2/\max(w,h)$.

To compute the subsidiary point correspondences, we reconstruct 3D points $\mathbf{X}_i$, project them into the two images, and distort the projected coordinates via (2). The canonical projection matrices for the two images can be parameterized as $\mathtt{P} = [\mathtt{I}|\mathbf{0}]$ and $\mathtt{P}' = [\mathtt{I}|\mathbf{e}]$ (Hartley and Zisserman, 2003, Result 10.2). If we denote the distortion function of (2) by $\delta(\mathbf{x},\lambda)$, we have $\hat{\mathbf{x}}_{u,i} = \mathtt{P}\mathbf{X}_i$, $\hat{\mathbf{x}}'_{u,i} = \mathtt{P}'\mathbf{X}_i$, $\hat{\mathbf{x}}_{d,i} = \delta(\hat{\mathbf{x}}_{u,i},\lambda)$, and $\hat{\mathbf{x}}'_{d,i} = \delta(\hat{\mathbf{x}}'_{u,i},\lambda)$.

We obtain the initial estimates of $\mathtt{F}$ and λ with the linear algorithm of Section 4. The initial estimates of the 3D points $\mathbf{X}_i$ are computed by triangulating the point correspondences after undistorting them with the initial estimate of λ. The 3D reconstruction could be computed using the optimal triangulation given in Hartley and Sturm (1997) and Hartley and Zisserman (2003, Section 12.5). However, for speed reasons it is advantageous to use the linear triangulation method of Hartley and Zisserman (2003, Section 12.2), which is more than twice as fast as the optimal method and provides the initial estimates with sufficient accuracy.

6. Reconstruction ambiguity

It is well known that the scene can be reconstructed up to an affine transformation (twelve parameters) if the camera has undergone a pure translation (Moons et al., 1993; Armstrong et al., 1994; Hartley and Zisserman, 2003, Section 10.4). In this section, we will show that under the assumptions that the division model implicitly makes (square pixels with zero skew and principal point at the center of the image; see Section 2), the reconstruction ambiguity is restricted even more: the scene can be reconstructed up to the usual similarity transformation ambiguity (seven parameters) that would be present if the interior orientation of the camera were known plus an unknown scale factor in the viewing direction (the direction of the optical axis). This has many practical implications for industrial applications. If, as is often the case, the camera is mounted over a conveyor belt such that the image plane is approximately parallel to the conveyor belt, a stereo reconstruction of the scene will preserve parallelism as well as orthogonality to the conveyor belt. This is important in inspection applications. In particular, this means that objects can be segmented very easily by thresholding the reconstructed points.

[2] As noted by Fitzgibbon (2001), this is similar to the high-curvature bias in some ellipse fitting algorithms.

To see that the reconstruction is ambiguous only up to a similarity transformation plus a scaling in the viewing direction, we note that the true projection matrices for our problem are given by $\tilde{\mathtt{P}} = \mathtt{K}[\mathtt{I} \mid \mathbf{0}]$ and $\tilde{\mathtt{P}}' = \mathtt{K}[\mathtt{I} \mid \mathbf{t}]$, where $\mathbf{t}$ is the translation between the cameras.[3] Note that this parameterization of the projection matrices implies that we have already factored out an ambiguity up to a rigid transformation. In the Gold Standard algorithm, we have parameterized the reconstruction by $\mathtt{P} = [\mathtt{I}|\mathbf{0}]$ and $\mathtt{P}' = [\mathtt{I}|\mathbf{e}]$. Thus, we seek a 3D homography $\mathtt{H}$ such that $\mathtt{P} = \tilde{\mathtt{P}}\mathtt{H}$ and $\mathtt{P}' = \tilde{\mathtt{P}}'\mathtt{H}$. It is easy to verify that $\mathtt{H}$ is given by

$$\mathtt{H} = \begin{bmatrix} \mathtt{K}^{-1} & \mathbf{0} \\ \mathbf{0}^{\top} & 1 \end{bmatrix} \tag{20}$$

since we have $\tilde{\mathtt{P}}\mathtt{H} = [\mathtt{I}|\mathbf{0}] = \mathtt{P}$ and $\tilde{\mathtt{P}}'\mathtt{H} = [\mathtt{I}|\mathtt{K}\mathbf{t}] = [\mathtt{I}|\mathbf{e}] = \mathtt{P}'$, i.e., $\mathbf{e} = \mathtt{K}\mathbf{t}$. The last equation only holds up to scale since we neither can reconstruct the true length of $\mathbf{t}$ nor that of $\mathbf{e}$. Thus, $s\mathbf{e} = \mathtt{K}\mathbf{t}$, where s is an undetermined scale factor. Under the above assumptions, the calibration matrix has the following form:

$$\mathtt{K} = \begin{bmatrix} c & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{21}$$

where c is the principal distance of the camera. Thus, the reconstruction ambiguity is actually a non-uniform scaling:

$$\mathtt{H} = \begin{bmatrix} s/c & 0 & 0 & 0 \\ 0 & s/c & 0 & 0 \\ 0 & 0 & s & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{22}$$

Since the overall scale s of the reconstruction cannot be determined, this means that the reconstruction is ambiguous only up to an unknown scaling of the z coordinates. The xy coordinates of the reconstruction will have the correct aspect ratio and there will be no skew in the reconstruction. Hence, the x and y axes will be orthonormal and the z axis will be orthogonal to the xy plane.

Finally, we note that if the principal point is not the center of the image, (20) implies that the reconstruction will have a skew in the xz and yz planes, i.e., the z axis will not be orthogonal to the xy plane. The farther the principal point is from the image center, the larger the skew will be, i.e., if the principal point is close to the image center the skew will be negligible. Furthermore, nonsquare pixels will lead to a non-orthonormal reconstruction in the xy coordinates. In addition, a non-zero skew in the calibration matrix leads to a skew in the xy plane of the reconstruction, i.e., the x axis will not be orthogonal to the y axis. Finally, all deviations from the assumed model (distortion center in the image center, zero skew, and square pixels) will lead to nonlinear distortions of the reconstruction. The closer the sensor geometry fulfills the assumptions, the smaller the nonlinear distortions will be. As the examples in Section 9.2 show, in practice the nonlinear distortions are negligible.

7. Degenerate configurations

From (13) and (14), we note that if the true value of λ is 0 and if the points $\mathbf{x}_d$ and $\mathbf{x}'_d$ are known without error, i.e., if they are undisturbed by noise, the matrix $\mathtt{A}$ in (15) becomes singular. Consequently, one eigenvalue μ will be ∞. The same situation will occur in (16). Note that this is an algorithmic degeneracy of the proposed minimal and overdetermined linear solvers. The Gold Standard solver does not have this degeneracy. A special treatment for this case could be defined easily for both solvers. However, since this special case mainly occurs in synthetic experiments and with probability zero in practice, we have not included any special treatment in our implementation of the minimal solver. For the overdetermined solver, this degeneracy is solved automatically by the extension of the algorithm described at the end of Section 4.

A more interesting degeneracy occurs if the true epipole is the center of the radial distortions, i.e., if $\mathbf{e} = (0,0,1)^{\top}$. From the discussion in Section 6 (see also Hartley and Zisserman, 2003, Eq. (9.3)), we can see that for pure camera translations we have $\mathbf{e} = \mathtt{K}\mathbf{t}$. Thus, under the assumptions of the division model, $\mathbf{e} = (0,0,1)^{\top}$ (up to scale, of course) if $\mathbf{t} = (0,0,t_z)^{\top}$, i.e., if the translation is exactly along the optical axis.

Suppose a 3D point $\mathbf{X} = (X,Y,Z)^{\top}$ is projected into the two images to the undistorted points $\mathbf{x}_u$ and $\mathbf{x}'_u$. From Hartley and Zisserman (2003, Eq. (9.6)), we can derive that $\mathbf{x}'_u = \mathbf{x}_u t_z/Z$, i.e., the points are related to each other by a pure scaling about the center of radial distortions. From (1) and (2), it is also clear that the division model implies a pure scaling about the center of radial distortions. This gives us an indication that it might not be possible to determine λ and $\mathtt{F}$ uniquely in this case. As the following theorem shows, this is indeed true.

Theorem 1. *Under the division model, for a pure camera translation along the optical axis, it is impossible to determine λ and $\mathtt{F}$ uniquely.*

To prove the theorem, we will show that for any configuration of 3D points $\mathbf{X}_{t,i}$, $i = 1,\ldots,n$, and any given values of the distortion coefficient λ_t, the principal distance c, and the translation $\mathbf{t} = (0,0,t_z)^{\top}$, there exists a configuration of 3D points $\mathbf{X}_{v,i}$ for $\lambda_v = 0$ that projects to the same image points in the two images with the same values of c and $\mathbf{t}$.[4]

With the above definitions, a 3D point $\mathbf{X} = (X,Y,Z)^{\top}$ projects to the following undistorted points in the two images:

$$\mathbf{x}_u = \frac{c}{Z}\begin{pmatrix} X \\ Y \end{pmatrix} \qquad \mathbf{x}'_u = \frac{c}{Z + t_z}\begin{pmatrix} X \\ Y \end{pmatrix} \tag{23}$$

The points will then be distorted by (2) with the respective value of λ.

To simplify the proof of the theorem, we make use of the above result that the perspective distortions caused by the translation along the optical axis and the radial distortions cause a pure radial scaling of the points in the image. This allows us to use polar coordinates $(r,\theta)^{\top}$ around the distortion center for the image coordinates and cylindrical coordinates $(R,\theta,Z)^{\top}$ around the optical axis for the 3D coordinates. Since 3D points with a particular value of θ project to image points with the same value of θ, we can ignore the actual value of θ. Thus, the 3D projections are given by:

$$r_u = \frac{cR}{Z} \qquad r'_u = \frac{cR}{Z + t_z} \tag{24}$$

In polar coordinates, (2) is given by:

$$r_d = \frac{2r_u}{1 + \sqrt{1 - 4\lambda\|r_u\|^2}} \tag{25}$$

To simplify the notation, we will drop the index variable i from the true 3D points $\mathbf{X}_t$, the virtual 3D points $\mathbf{X}_v$, and their respective projections. Thus, to prove the theorem, we will show that for the true 3D points $(R_t,Z_t)^{\top}$ that are projected by the true distortion

[3] Since we are performing an uncalibrated reconstruction, we do not know the values of $\mathtt{K}$ and $\mathbf{t}$, of course.

[4] Of course, the 3D points must be chosen in such a way that the projection and distortion are well defined under the selected value of λ. In particular, the square root in the denominator of (2) must be nonnegative.

coefficient λ_t to the true image points $r_{d,t}$ and $r'_{d,t}$, there exist virtual 3D points $(R_v,Z_v)^\top$ that are projected by the virtual distortion coefficient $\lambda_v = 0$ to the image points $r_{d,v}$ and $r'_{d,v}$ such that

$$r_{d,t} = r_{d,v} \quad \text{and} \quad r'_{d,t} = r'_{d,v} \tag{26}$$

By regarding c, t_z, λ_t, $(R_t,Z_t)^\top$, and λ_v as fixed, (26) defines two nonlinear equations for the two unknowns $(R_v,Z_v)^\top$. These equations can be solved in closed form. Since the derivation of the solution is quite tedious, we will only present the final result:

$$\begin{pmatrix} R_v \\ Z_v \end{pmatrix} = \frac{t_z}{d} \begin{pmatrix} 2R_t \\ fZ_t \end{pmatrix} \tag{27}$$

where

$$\begin{aligned} d &= t_z f' + Z_t(f' - f) \\ f &= 1 + \sqrt{1 - 4\lambda_t \frac{c^2 R_t^2}{Z_t^2}} \\ f' &= 1 + \sqrt{1 - 4\lambda_t \frac{c^2 R_t^2}{(Z_t + t_z)^2}} \end{aligned} \tag{28}$$

Corollary 2. *Unter the division model, for a pure camera translation along the optical axis, there are infinitely many possible solutions for λ and $\mathtt{F}$.*

As the proof of Theorem 1 shows, there is a bijection between the 3D points for λ_r and $\lambda_v = 0$. We can apply the bijection in reverse from $\lambda_v = 0$ to any desired new value of λ. Therefore, for each arbitrarily chosen λ, there exists a valid 3D reconstruction.

Conjecture 3. *For any distortion model that only models radial distortions around the optical axis, there are infinitely many possible solutions for the radial distortion parameters for a pure camera translation along the optical axis.*

This is plausible since the derivation of (27) only makes use of the fact that the projection and distortion of the 3D points are unique and well defined. Unlike for the division model, for other radial distortion models it might not be possible to find a closed-form solution for the reconstruction ambiguity.

In practice, Theorem 1 implies that the closer the camera translation is to being along the optical axis, the more uncertain the determination of λ and $\mathtt{F}$ become. We will verify this conclusion experimentally in Section 9.1. Note that since the translation along the optical axis is a structure degeneracy, all uncalibrated stereo algorithms that take radial distortions into account have this property.

Finally, we mention another degeneracy that does not seem to have been described explicitly in the literature so far, presumably because it is trivial to derive. Suppose the camera is an affine camera, i.e., $c = \infty$. An affine camera can easily be realized by means of a telecentric lens (Hornberg, 2006, Section 4.2.14; Steger et al., 2008, Section 2.2.4). It is obvious that for affine cameras under pure camera translations all rays in the two images are parallel to each other, i.e., they intersect in the same point on the plane at infinity. Therefore, information about 3D cannot be determined in this case, no matter whether radial distortions are taken into account or not. In practice, this means that stereo reconstruction is impossible with telecentric lenses under pure camera translations. For standard perspective lenses, this implies that as the principal distance of the camera becomes longer, the determination of $\mathtt{F}$ becomes more uncertain under pure camera translations.

8. Automatic determination of the fundamental matrix and the radial distortion

To determine the fundamental matrix $\mathtt{F}$ and the radial distortion coefficient λ automatically, we use an appropriately modified version of the RANSAC algorithm described in Hartley and Zisserman (2003, Algorithm 11.4). For convenience, we will briefly describe this algorithm and the modifications that are necessary to make it work properly for our problem.

The algorithm first extracts interest points in the two images. It then computes a set of putative point matches based on the proximity of the point pairs and the similarities of their intensity neighborhoods. Next, a RANSAC algorithm is executed to select a maximally consistent set of inliers. The RANSAC algorithm selects a random sample of three correspondences and computes $\mathtt{F}$ and λ with the minimal solver described in Section 3. For all solutions returned by the minimal solver, it computes the distance d for all putative point correspondences and counts the number of inliers for which the distance is below a threshold t. The number of iterations of the RANSAC algorithm is determined adaptively based on the maximum set of inliers that the algorithm has found so far. At the end of the RANSAC algorithm, the $\mathtt{F}$ and λ with the largest number of inliers are selected. Next, $\mathtt{F}$ and λ are re-estimated either with the overdetermined linear solver of Section 4 or the Gold Standard solver of Section 5. After this, improved point correspondences are obtained by matching points within a suitably restricted epipolar band (guided matching) and running the RANSAC algorithm on the resulting putative point correspondences. The final estimates of $\mathtt{F}$ and λ are then obtained with the improved correspondences either with the overdetermined linear solver or the Gold Standard solver.

In the algorithm above, the only modification to Hartley and Zisserman (2003, Algorithm 11.4) so far has been to replace the 7-point algorithm with the minimal solver of Section 3. However, further modifications are necessary to make the algorithm robust. We will first take a look at how the distance d should be defined. This distance measure already was discussed to some extent in Section 4. The basic constraint that the point correspondences must fulfill is that the epipolar distance of the point matches must be small. In the context of the algorithm in this section, this means that the distance of a point to its corresponding epipolar line must be smaller than the threshold t. This condition must be fulfilled in both images simultaneously. As discussed in Section 4, one method to compute the epipolar distance is to undistort the point coordinates and to measure the epipolar distance in the undistorted coordinate system. However, this leads to the fact that extreme values of λ, which almost assuredly will be returned by the minimal solver, will cause artificially small distance values to be generated. Therefore, to make the algorithm robust, we must use the second algorithm described in Section 4, where the epipolar distance is computed in the original, distorted image coordinate system by determining the closest point to the respective point on the corresponding epipolar line, to distort that point, and to measure the distance between the two distorted points.

Fig. 1 shows two typical examples that highlight the difference between the two algorithms to compute the distance d and its effect on the number of inliers. In this example, a set of 1000 point correspondences was generated synthetically by creating 1000 random 3D points that were restricted to a suitable 3D box and by projecting them into two images (which, conceptually, are of dimension 640 × 480). The synthetic camera had a principal distance of 500 (pixels). The projected points were then distorted with $\lambda = -10^{-6}$ (1/pixel^2). Next, Gaussian noise with standard deviation $\sigma = 1$ (pixel) was added to the point coordinates. The graphs in Figs. 1a and c were generated using one realization of this pro-

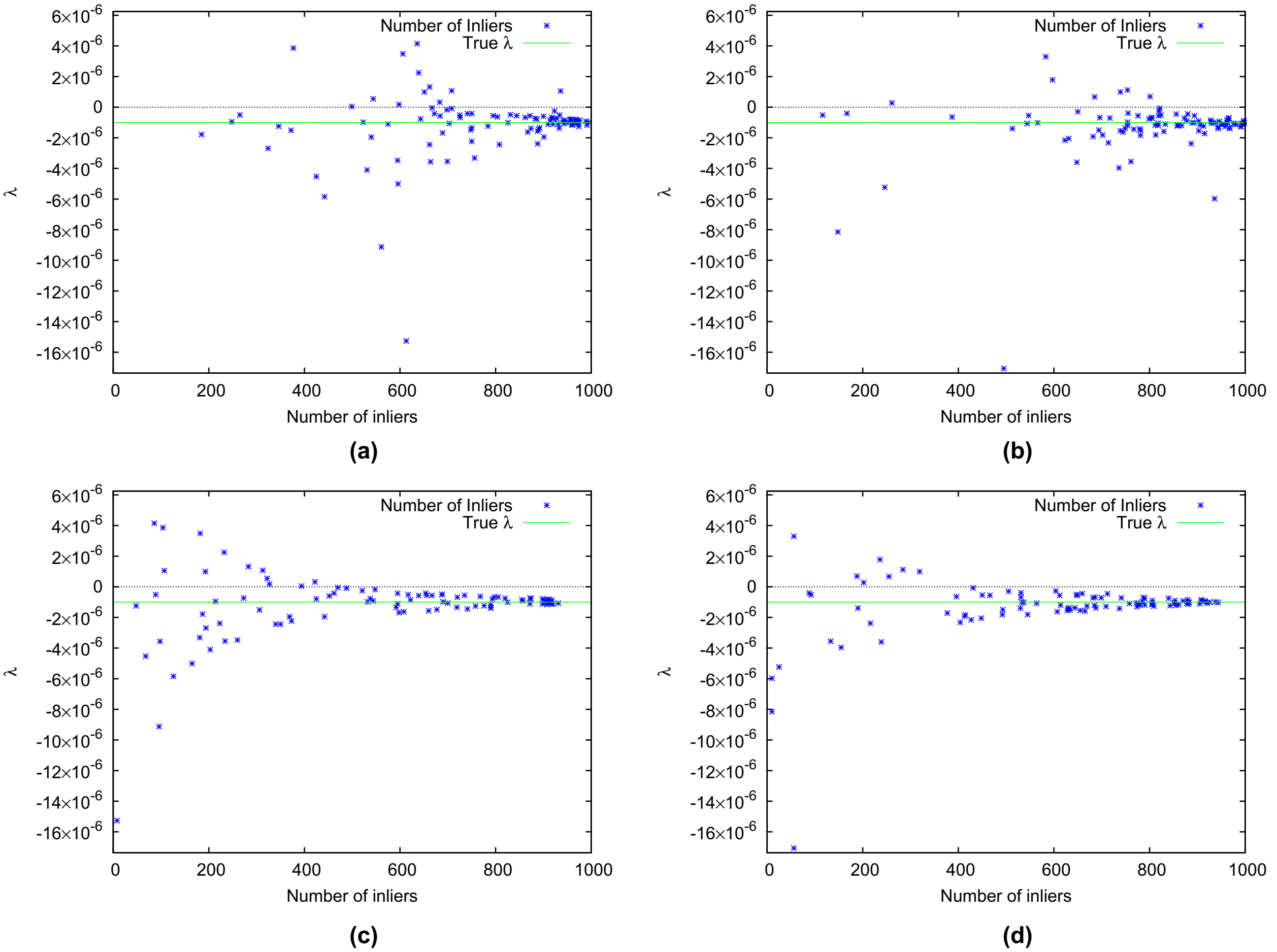


Fig. 1. Comparison of the different distance measures used for counting inliers in the RANSAC algorithm (see text for details). (a) and (b) show the number of inliers if the epipolar distance is computed after undistorting the points. (c) and (d) show the number of inliers if the epipolar distance is computed in the original image space using the distorted points with the same data sets as in (a) and (b), respectively. Note the completely erroneous values for λ in (a) and (b) that receive significantly more than 900 inliers and the wide distribution of values of λ that receive a large number of inliers. In contrast, note that the distribution of the values of λ that receive a large number of inliers is concentrated around the true value of λ in (c) and (d).

cess, while Figs. 1b and d were generated using a second realization of this process. Note that the two data sets consist entirely of inliers. The RANSAC algorithm was executed with 100 trials, i.e., minimal solutions were generated for 100 sets of three randomly selected point correspondences, and the number of inliers were counted using (17) and (18) (with $n = 1$) to compute the distance for each of the 1000 point correspondences for the value of $\mathtt{F}$ and λ returned by the minimal solver. To have a reasonable chance of obtaining the entire set of 1000 inliers, the distance threshold was set to $t = 3\sigma$.

Figs. 1a and b show the result of counting inliers when the distances are computed in the undistorted image coordinate system, while Figs. 1c and d show the results of computing the distances in the distorted image coordinate system. All graphs show the number of inliers for each of the 100 trials of the RANSAC algorithm as well as the value of λ (within the bounds for λ given by (4) and (6)) returned by the minimal solver in the respective trial. In Fig. 1a, we can see an inlier set with 936 inliers for $\lambda = 1.06 \times 10^{-6}$ (1/pixel^2), i.e., for a distortion that has the wrong sign, but approximately the correct magnitude. Similarly, in Fig. 1b, we can see an inlier set that also has 936 inliers for $\lambda = -5.97 \times 10^{-6}$ (1/pixel^2), i.e., for a distortion that is off by a factor of 6. We can also see that the number of inliers is very large for a substantial range of λ, i.e., the distribution of large numbers of inliers is very broad around the true value of λ. Thus, it can be seen that there is a substantial danger that the RANSAC algorithm will select incorrect values for $\mathtt{F}$ and λ if the distances are computed in the undistorted image coordinate system. In contrast, Figs. 1c and d show that the distribution of large numbers of inliers is quite narrow around the true value of λ if the distances are computed in the distorted image coordinate system. In particular, the two extreme cases mentioned above now only have 107 and 10 inliers, respectively. Therefore, we can see that with this method of computing the distances, it will almost be assured that the RANSAC algorithm finds the correct values of $\mathtt{F}$ and λ. To make the algorithm consistent, the guided matching also uses this distance measure to calculate the epipolar bands.

The above examples also illustrate a fact that has been noted frequently, e.g., in Scherer-Negenborn and Schaefer (2010): the standard method of computing the number of iterations of the RANSAC algorithm is too optimistic because it does not take into account that the points will be contaminated by noise. Note that in the above examples all 1000 points are inliers. Yet, the RANSAC algorithm produces inlier sets with anywhere from 3 to 1000 inliers. This is caused by the fact that some of the samples will have the three points clustered together in one area of the image. In these cases, the transformation is, in effect, extrapolated across the entire image, and therefore necessarily inaccurate. This problem is solved elegantly in Scherer-Negenborn and Schaefer (2010) by adding a term to the standard estimate of the number of iterations that takes into account that the points that are used to create the minimum solution are spread well across the image. We therefore use the formula of Scherer-Negenborn and Schaefer (2010) to estimate the number of samples required by the RANSAC algorithm adaptively.

We also note that Raguram et al. (2009) have recently proposed an algorithm that takes the uncertainties of the point coordinates as well as the uncertainty of the minimal solution of the fundamental matrix for the general motion case without distortions into account within the RANSAC algorithm. In future research, it might be useful to determine whether the approach proposed by Raguram et al. (2009) would also be beneficial for the case we consider in this paper.

9. Performance evaluation

In Section 9.1, we will study the accuracy of the results that can be achieved with the overdetermined linear solver described in Section 4 and the Gold Standard solver described in Section 5. Since an accuracy evaluation requires the ground truth to be available, the accuracy evaluation will be performed on synthetic data. As noted in Section 7, there are some degeneracies that may affect the accuracy of the results in practice. Therefore, another goal of Section 9.1 is to examine how these degeneracies affect the results in practice. Finally, we also compare the overdetermined linear solver for pure camera translations to the overdetermined linear solver for general camera motions proposed by Fitzgibbon (2001) to determine whether taking into account the actual type of camera motion improves the results.

In Section 9.2, we will provide examples on real data that demonstrate that the automatic determination of the fundamental matrix and the radial distortion described in Section 8 works reliably. This inherently also tests the minimal solver described in Section 3. Furthermore, we will show that the reconstruction ambiguity derived in Section 6 (a similarity transformation plus a scaling in the viewing direction) can actually be achieved in practice. In addition, we will show that severe reconstruction errors can result if radial distortions are not taken into account. Finally, we will show that using minimal and overdetermined solvers for the case of general camera motion does not provide adequate results for several reasons.

9.1. Synthetic data

9.1.1. Accuracy

We first study the accuracy that can be achieved with the overdetermined linear solver of Section 4 and the Gold Standard solver of Section 5.

To determine the accuracy, we are primarily interested in two quantities: the accuracy of the estimated radial distortion coefficient λ and the residual error that can be achieved for different true values of λ and for different amounts of noise by which the image coordinates are disturbed. On the one hand, the accuracy is measured as the deviation of the extracted value of λ from its true value. Ideally, this should be zero. On the other hand, the residual errors allow us to assess the accuracy of the computed fundamental matrix indirectly. If the extracted λ and $\mathtt{F}$ are correct, we expect the residual errors to have the same magnitude as the noise that was added to the image coordinates. Therefore, substantial deviations from this expectation allow us to infer that $\mathtt{F}$ is erroneous even if λ has a comparatively small error.

In the experiments, we assume images of size 640 × 480. From (4) and (6), it can be seen that for this image size we must have $-1.736 \times 10^{-5} \leqslant \lambda \leqslant 6.25 \times 10^{-6}$ (1/pixel^2). We did not use the entire range of possible distortions for the experiments, but used a smaller range of $-6 \times 10^{-6} \leqslant \lambda \leqslant 4 \times 10^{-6}$ (1/pixel^2). The discussion below will show why this choice is justified. A set of 250 points was created randomly within a suitable 3D box in such a way that the projections of the 3D points lay completely within the 640 × 480 images and filled them reasonably well after the distortion with a particular λ. The projection was performed with a synthetic camera that had a principal distance of 500 (pixels). The 3D translation between the two cameras was chosen as $(4, 1, 3)^\top$ (mm). This 3D translation was used throughout the experiments. The projected points were then distorted with different values of λ. Gaussian noise with varying standard deviation $0 \leqslant \sigma \leqslant 2$ (pixels) was then added to the projected points. This was repeated 100 times for every value of λ and σ and the median values of the estimated λ and residual error, as well as the 10% and 90% quantiles of these two values were determined.

Fig. 2 displays the result of this accuracy evaluation. Fig. 2a shows the error of the median estimated λ plotted for $\lambda \in [-6 \times 10^{-6}, 4 \times 10^{-6}]$ (1/pixel^2) and $\sigma \in [0, 2]$ (pixels) for the overdetermined linear solver. Please note that the error of the median estimated λ is the same as the median of the errors of the estimated values of λ. Consequently, we will refer to this error as the median bias of λ. As described at the end of Section 4, the overdetermined linear solver switches to a solver for the pure translation case without radial distortions if it is unable to find a valid solution. This happens for very small values of λ, in particular for $\lambda = 0$. To understand the behavior of the overdetermined linear solver with radial distortions without being distracted by the results of the solver without radial distortions, the value $\lambda = 0$ was excluded from the experiments for this solver. It can be seen that the results of the overdetermined linear solver are relatively unbiased except for small absolute values of λ and large amounts of noise, which cause a significant bias of the estimated λ. As was also observed in Fitzgibbon (2001), the bias for small λ and large σ is towards more extreme distortions of the same sign. Note, however, that the results of the overdetermined linear solver become less biased for larger distortions. Nevertheless, for larger values of λ, large amounts of noise still introduce some bias. For example, the estimated distortion for $\lambda = -6 \times 10^{-6}$ (1/pixel^2) and $\sigma = 2$ (pixels) is biased by approximately 7%. These results show that the above choice to restrict the tests to $\lambda \in [-6 \times 10^{-6}, 4 \times 10^{-6}]$ (1/pixel^2) is reasonable because the bias only becomes smaller for larger absolute values of λ.

Fig. 2b shows the median bias of the estimated values of λ for the Gold Standard solver (including $\lambda = 0$). It can be seen that the Gold Standard solver returns unbiased results that are very accurate. The results are up to two orders of magnitude more accurate than those of the overdetermined linear solver. Furthermore, we can see that, although the Gold Standard solver is initialized with the potentially biased results of the overdetermined linear solver, it has converged reliably to the correct solution.

9.1.2. Precision

For the Gold Standard solver, it is also interesting to use the interdecile range that we have computed as a measure of the precision of the results. We would expect the precision of λ to be proportional to σ and to be inversely proportional to the square root of the number of point correspondences n. We do not show the corresponding graphs here since the results are easy to summarize: the interdecile range does not depend on the value of λ and depends linearly on σ. For $\sigma = 2$ (pixels), the interdecile range of the estimated values of λ is approximately 2×10^{-7} (1/pixel^2). Thus, we can see that the expectation that the precision of λ is proportional to σ is fulfilled in practice. To verify the dependence of the precision on $1/\sqrt{n}$, an experiment was performed in which

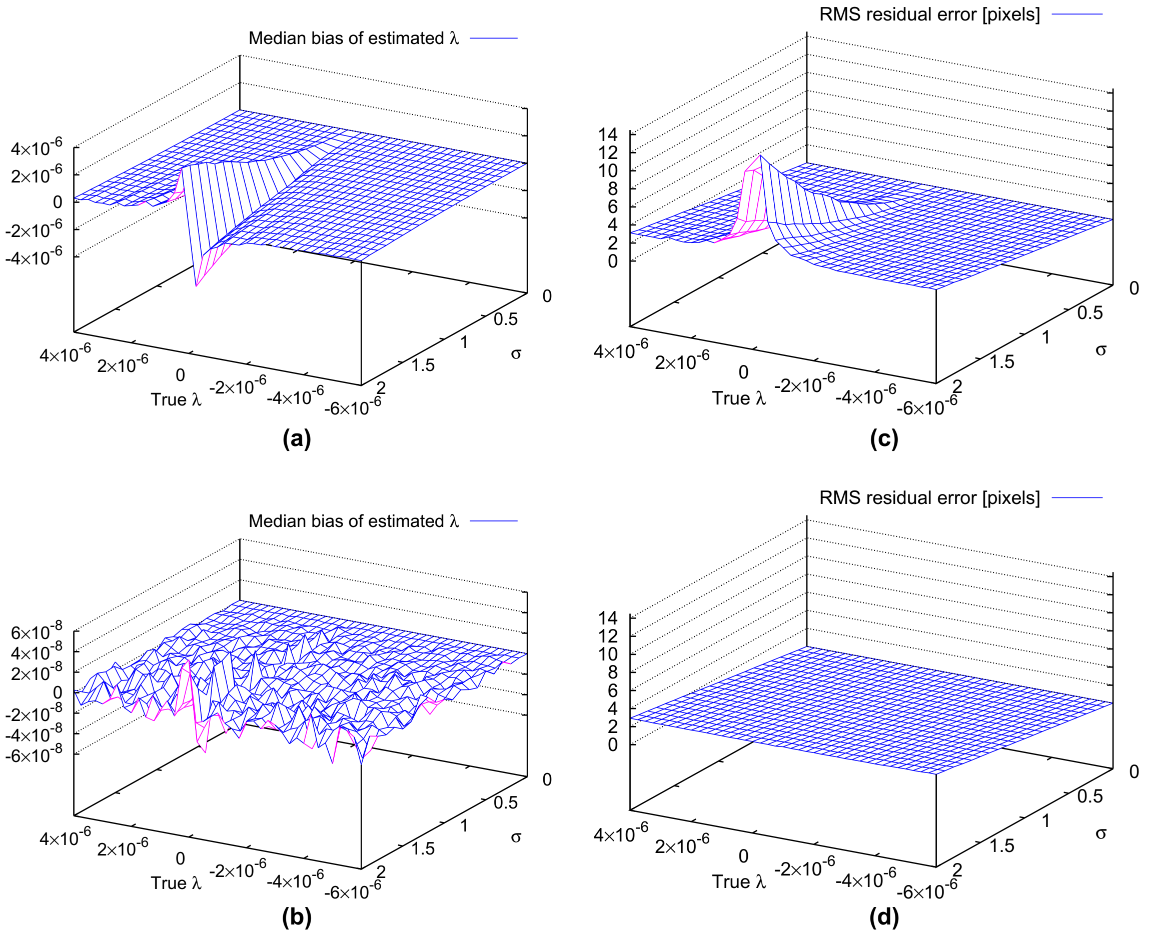


Fig. 2. (a) Median bias of radial distortion coefficient λ estimated by the overdetermined linear solver for $\lambda \in [-6 \times 10^{-6}, 4 \times 10^{-6}]$ (1/pixel^2) and Gaussian noise added to the point correspondences with standard deviation $\sigma \in [0, 2]$ (pixels). (b) Median bias of radial distortion coefficient λ estimated by the Gold Standard solver. (c) RMS residual error of the solutions returned by the overdetermined linear solver (pixels). (d) RMS residual error of the solutions returned by the Gold Standard solver (pixels).

the number of point correspondences was varied from 10 to 250. This experiment showed that for the geometry described above, the interdecile range of the estimated values of λ is approximately $1 \times 10^{-6}/\sqrt{n}$ (1/pixel^2) for $\sigma = 0.5$ (pixels). Therefore, we can see that the parameter λ can be determined very precisely with the Gold Standard solver.

9.1.3. Residual errors

Since the overdetermined linear solver returns biased results, while the Gold Standard solver returns unbiased results, it is interesting to examine the residual errors that result from the different estimates. In the evaluation, we determine the residual error by (18), then divide by $2n$, where n is the number of point correspondences, and compute the square root. This is the root-mean-square (RMS) residual error per point, i.e., a measure for the mean distance of a distorted point from its closest distorted point on the corresponding epipolar line. The RMS residual errors are shown in Figs. 2c and d. We can see that the residual errors are quite similar for the overdetermined linear solver and the Gold Standard solver, except for the case of small absolute values of λ and large σ. In particular, the remaining bias for large absolute values of λ and large values of σ results in a very small increase of the RMS residual error (for example, less than 5% for $\lambda = -6 \times 10^{-6}$ (1/pixel^2) and $\sigma = 2$ (pixels)).

As noted above, ideally we would expect the RMS residual errors to be the same size as the noise that we have added to the points $\left(\sqrt{2}\sigma\right)$. As can be seen from Fig. 2d, this expectation is fulfilled almost exactly for the Gold Standard solver.

9.1.4. Accuracy for realistic noise levels

The above results need further examination. In particular, the bias of the overdetermined linear solver merits further evaluation. In the above experiments, a very large range of noise standard deviations was used. Real subpixel-precise interest point detectors, such as the Harris detector (Harris and Stephens, 1988) or the Förstner detector (Förstner, 1994) are typically able to extract the points with a much greater accuracy. We observe accuracies of 0.1–0.25 pixels with these detectors in real applications. Therefore, the above experiments were repeated with a more restricted range of $\lambda \in [-10^{-6}, 10^{-6}]$ (1/pixel^2) and $\sigma \in [0, 0.5]$ (pixels) to obtain a better understanding of the performance of the two algorithms that can be expected in a real application if the true value of λ is small. The results of this evaluation are shown in Figs. 3a and b. The bias of the overdetermined linear solver can still be observed. However, as expected, it is much smaller. Again, the Gold Standard solver returns unbiased results. Figs. 3c and d show that the residual errors of the two algorithms are quite similar, except for very small absolute values of λ. Thus, we can see that for sufficiently large absolute values of λ and for realistic noise levels, the overdetermined linear solver is a viable alternative to the Gold Standard solver if speed is important in the application. On the other hand, if the highest possible accuracy is required, the Gold Standard algorithm should be used.

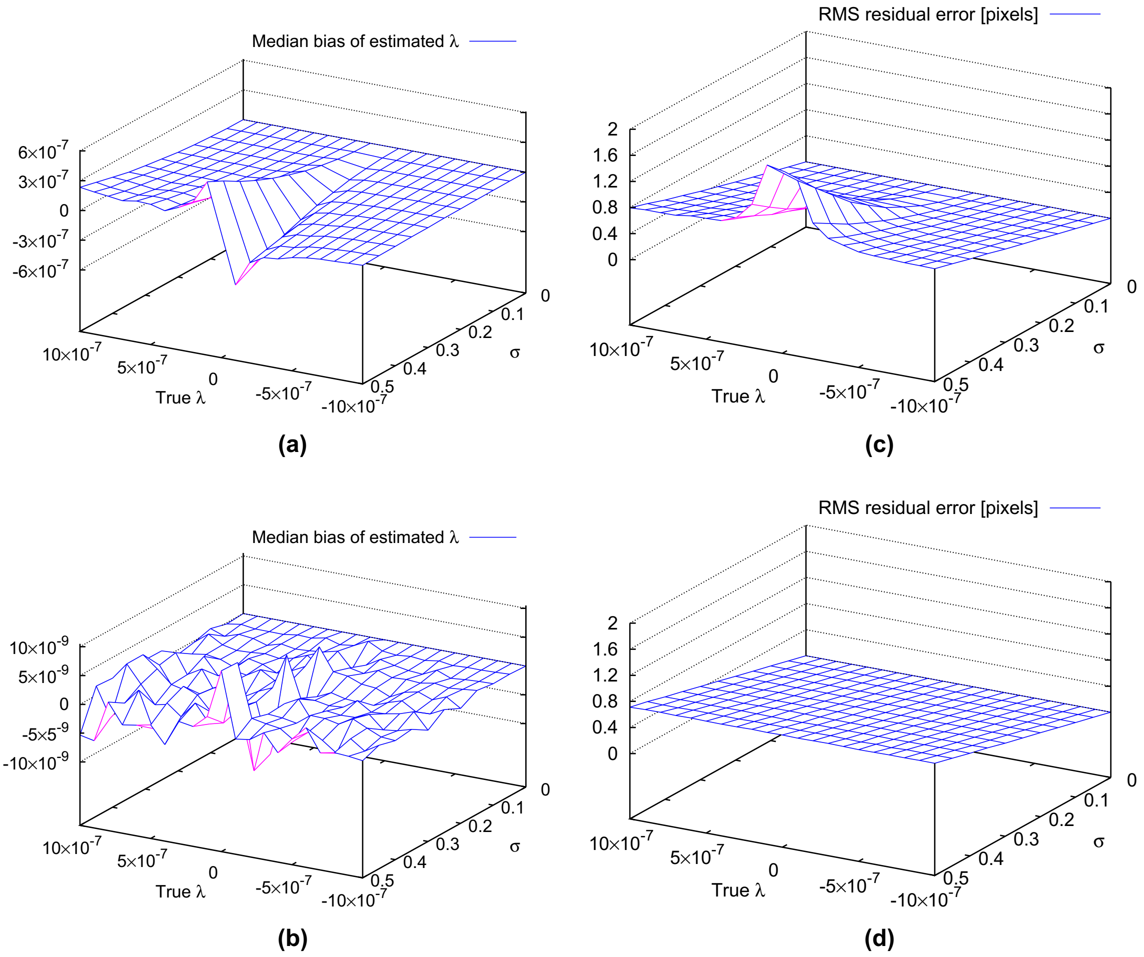


Fig. 3. (a) Median bias of radial distortion coefficient λ estimated by the overdetermined linear solver for $\lambda \in [-10^{-6}, 10^{-6}]$ (1/pixel^2) and Gaussian noise added to the point correspondences with standard deviation $\sigma \in [0, 0.5]$ (pixels). (b) Median bias of radial distortion coefficient λ estimated by the Gold Standard solver. (c) RMS residual error of the solutions returned by the overdetermined linear solver (pixels). (d) RMS residual error of the solutions returned by the Gold Standard solver (pixels).

9.1.5. Comparison to an overdetermined linear solver for general camera motions

We extend the evaluation by comparing our overdetermined linear solver for pure camera translations to the overdetermined linear solver for the case of general camera motions proposed by Fitzgibbon (2001). The results displayed in Fig. 4 show that the overdetermined linear solver for general camera motions returns extremely biased results. Because of this, we had to restrict the noise to the range $\sigma \in [0, 1]$ (pixels). Even in this range, the results are so biased that frequently no solutions within the allowable range for λ given by (4) and (6) are found. In this case, our implementation internally switches to a linear solver that does not consider radial distortions, analogously to the procedure described at the end of Section 4. This results in the tilted plane in the central part of the error plot in Fig. 4b. In contrast, the bias returned by the overdetermined linear solver for pure camera translations is smaller by an order of magnitude. These results show that the overdetermined linear solver for general camera motions can only be used if the noise in the point coordinates is extremely small and if the distortion is quite large. In contrast, the overdetermined linear solver for pure camera translations returns meaningful results for almost all relevant noise levels and distortions. In particular, it can be seen that it is no option to use the general overdetermined linear solver for the case of pure camera translations.

9.1.6. Behavior for the structure degeneracy

We now turn our attention to the structure degeneracy described in Section 7. As noted there, if the camera translation is purely along the optical axis it is impossible to determine λ and $\mathbb{F}$ uniquely. To test the effects of this degeneracy in practice, translations of the form $\mathbf{t} = (t_x, 0, 3)^\top$ (mm) were created. The translation t_x was varied in the range $[-2, 2]$ (mm). Obviously, for $t_x = 0$ (mm) we obtain a pure translation along the optical axis. To be able to obtain plots that can be displayed, a realistic value of $\sigma = 0.25$ (pixels) was used.[5] The other parameters were identical as in the previous experiment. As above, the value $\lambda = 0$ was excluded for the overdetermined linear solvers, but was included for the Gold Standard solver. The results of this experiment are shown in Fig. 5.

As can be seen from Fig. 5a, the median bias of the overdetermined linear solver becomes progressively larger as t_x approaches 0. Consequently, there is a range of translations close to $t_x = 0$ (mm) for which the overdetermined linear solver is not able to determine a valid solution and switches to the overdetermined linear solver without radial distortions internally. This explains the tilted plane in the central part of the plot. Fig. 5d shows that the RMS residual errors actually become smaller by neglecting the radial distortions

[5] Testing $\sigma = 0$ (pixels) would only have shown that the structure degeneracy at $t_x = 0$ (mm) is reproduced exactly by the algorithms.

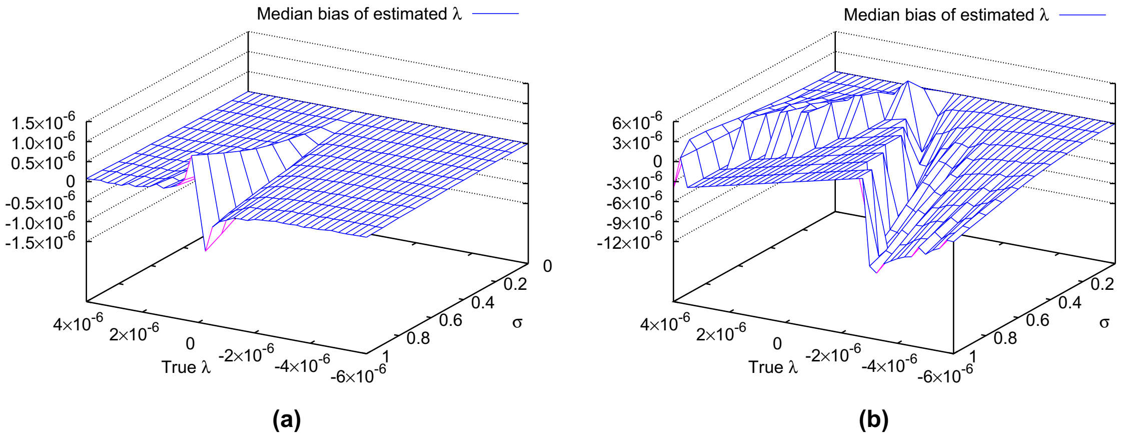


Fig. 4. (a) Median bias of radial distortion coefficient λ estimated by the overdetermined linear solver for pure camera translations for $\lambda \in [-10^{-6}, 10^{-6}]$ (1/pixel^2) and Gaussian noise added to the point correspondences with standard deviation $\sigma \in [0, 1]$ (pixels). (b) Median bias of radial distortion coefficient λ estimated by the overdetermined linear solver for general camera motions.

for small values of t_x. As above, the degeneracy for small values of λ can also be seen in these plots.

In contrast, Fig. 5b shows that the Gold Standard solver returns unbiased results for most values of λ and t_x. Since the Gold Standard solver is initialized with the results of the overdetermined linear solver, we can observe the expected result that $\lambda = 0$ is returned for $t_x = 0$ (mm). However, it can also be seen that for very small values of t_x (smaller than 0.2 mm) the results of the Gold Standard solver are slightly biased. Nevertheless, the small and approximately constant RMS residual errors in Fig. 5e demonstrate that the Gold Standard solver returns valid solutions in all cases.

Finally, Fig. 5c displays the results of the overdetermined linear solver for the general motion case. At first glance, it might seem that this solver returns better results than the overdetermined linear solver for the pure translation case since the tilted plane in Fig. 5a is mostly absent in Fig. 5c. Surprisingly, the overdetermined linear solver for general camera motions seems to be able to determine λ for $t_x = 0$ (mm) without switching to the fallback without distortions internally. As can be seen from Fig. 5f, this comes at the expense of large RMS residual errors for small values of t_x. Thus, Fig. 5f shows that the fundamental matrix determined by the overdetermined linear solver for general camera motions contains large errors. In addition, by comparing Fig. 5a and c, it can be seen that the general motion case results in significantly larger bias than the pure translation case. Again, it can be seen that the general camera motion solver should not be used for pure camera translations.

9.2. Real data

9.2.1. Setup

To test the automatic determination of the fundamental matrix and the radial distortion, a representative test set for an industrial application was used. Images of a printed circuit board (PCB) were acquired with an industrial camera (an IDS uEye UI-2240-M, which has a 1/2″′ chip with pixel size 4.65 μm and a resolution of 1280 × 1024 pixels) and a Pentax 4.8 mm lens with a 1 mm extension tube. Two sets of images were acquired. In the first set, the camera was mounted approximately perpendicular to the PCB, while in the second set the camera was mounted at an angle to the PCB. The PCB was moved along one of its edges in the world. Figs. 6a and d as well as 7a and d display the two test sets.

9.2.2. Automatic determination of the fundamental matrix and the radial distortion

The point matches obtained by running the algorithm of Section 8 with the Gold Standard solver on the images in Figs. 6a and d are shown in Figs. 6b and e. The interest points were extracted with the Förstner detector (Förstner, 1994). The algorithm found 1499 point matches with the distance threshold set to $t = 1$. The estimated radial distortion was $\lambda = -2.109 \times 10^{-7}$ (1/pixel^2). The RMS residual error per point correspondence was 0.3404 (pixels). This corresponds to a point accuracy of approximately 0.24 (pixels). If the overdetermined linear solver is used instead, the estimated radial distortion is $\lambda = -2.128 \times 10^{-7}$ (1/pixel^2) with an RMS residual error per point correspondence of 0.3438 (pixels). Thus, in this example both algorithms return only marginally different results. The two estimates of λ result in a difference of less than one pixel in the geometry of the undistorted images at the corners of the image. Thus, in this example, the overdetermined linear solver is a very good alternative to the Gold Standard solver, especially since the algorithm of Section 8 runs more than twice as fast if it is used.

The results of computing the fundamental matrix and radial distortion coefficient for the test set in which the camera was tilted are shown in Fig. 7. The algorithm of Section 8 found 1241 point matches with $t = 1$, which are shown in Figs. 7b and e. The Gold Standard solver estimated $\lambda = -2.108 \times 10^{-7}$ (1/pixel^2) with an RMS residual error per point correspondence of 0.3778 (pixels), while the overdetermined linear solver estimated $\lambda = -2.127 \times 10^{-7}$ (1/pixel^2) with an RMS residual error per point correspondence of 0.3780 (pixels). As can be seen, the results are very consistent with those of the first test set.

9.2.3. Stereo reconstruction

To test the reconstruction quality in a real application and to show that the reconstruction ambiguity described in Section 6 can be achieved in practice, a stereo reconstruction was performed for both test sets. Since the setup is uncalibrated, the image disparities are used to assess the quality of the stereo reconstruction.

For both test sets, the estimated λ was used to undistort the images. Furthermore, the images were projectively rectified using the algorithm described by Gluckman and Nayar (2001). The results of the undistortion and projective rectification for the first test set are shown in Figs. 6c and f, while the results for the second test set are shown in Figs. 7c and f.

Next, the variational stereo algorithm described by Drost (2008) was used to compute the disparities of each rectified stereo image pair.

Fig. 6g displays the resulting disparities in the areas for which the rectified images overlap sufficiently. As discussed in Section 5, the disparities represent an affine reconstruction of the scene up to an unknown scaling in distance. Thus, the true dis-

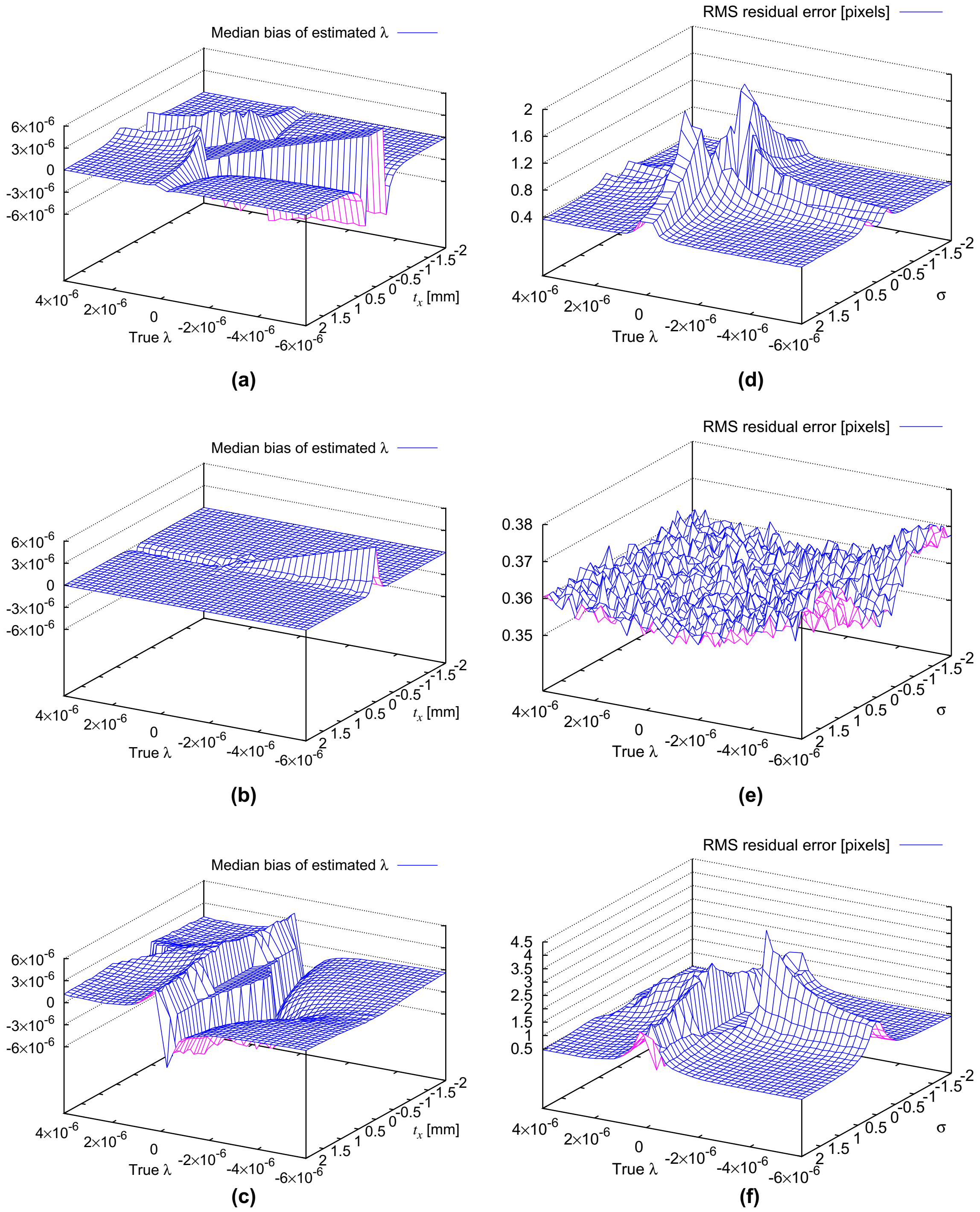


Fig. 5. (a) Median bias of radial distortion coefficient λ estimated by the overdetermined linear solver for pure camera translations for $\lambda \in [-6 \times 10^{-6}, 4 \times 10^{-6}]$ (1/pixel^2) and camera motions with $t_x \in [-2, 2]$ (mm), $t_y = 0$ (mm), and $t_z = 3$ (mm). (b) Median bias of radial distortion coefficient λ estimated by the Gold Standard solver for pure camera translations. (c) Median bias of radial distortion coefficient λ estimated by the overdetermined linear solver for general camera motions. (d) RMS residual error (pixels) of the solutions returned by the overdetermined linear solver for pure camera translations. (e) RMS residual error (pixels) of the solutions returned by the Gold Standard solver for pure camera translations. (f) RMS residual error (pixels) of the solutions returned by the overdetermined linear solver for general camera motions.

tances cannot be determined. However, the disparity is monotonically related to the distance. Therefore, Fig. 6h displays the disparities as a 3D plot, while Fig. 6i overlays the texture of Fig. 6c onto the 3D surface. As can be seen, the reconstructions of the PCB itself and the large IC on the PCB are perfectly planar. In addition, all the components are clearly discernible in the disparities. Therefore, it would be very easy to detect the components on the PCB by simple thresholding.

Fig. 7g displays the resulting disparities in the areas for which the rectified images overlap sufficiently for the second test set, while Figs. 7h and i display them as 3D plots. Again, it can be seen that the reconstruction of the PCB and the large IC on the PCB is perfectly flat (apart from a slight reconstruction error at the far side of the IC which is caused by the fact that the IC lies too close to the border of the rectified images). As before, the components on the PCB can be extracted easily.

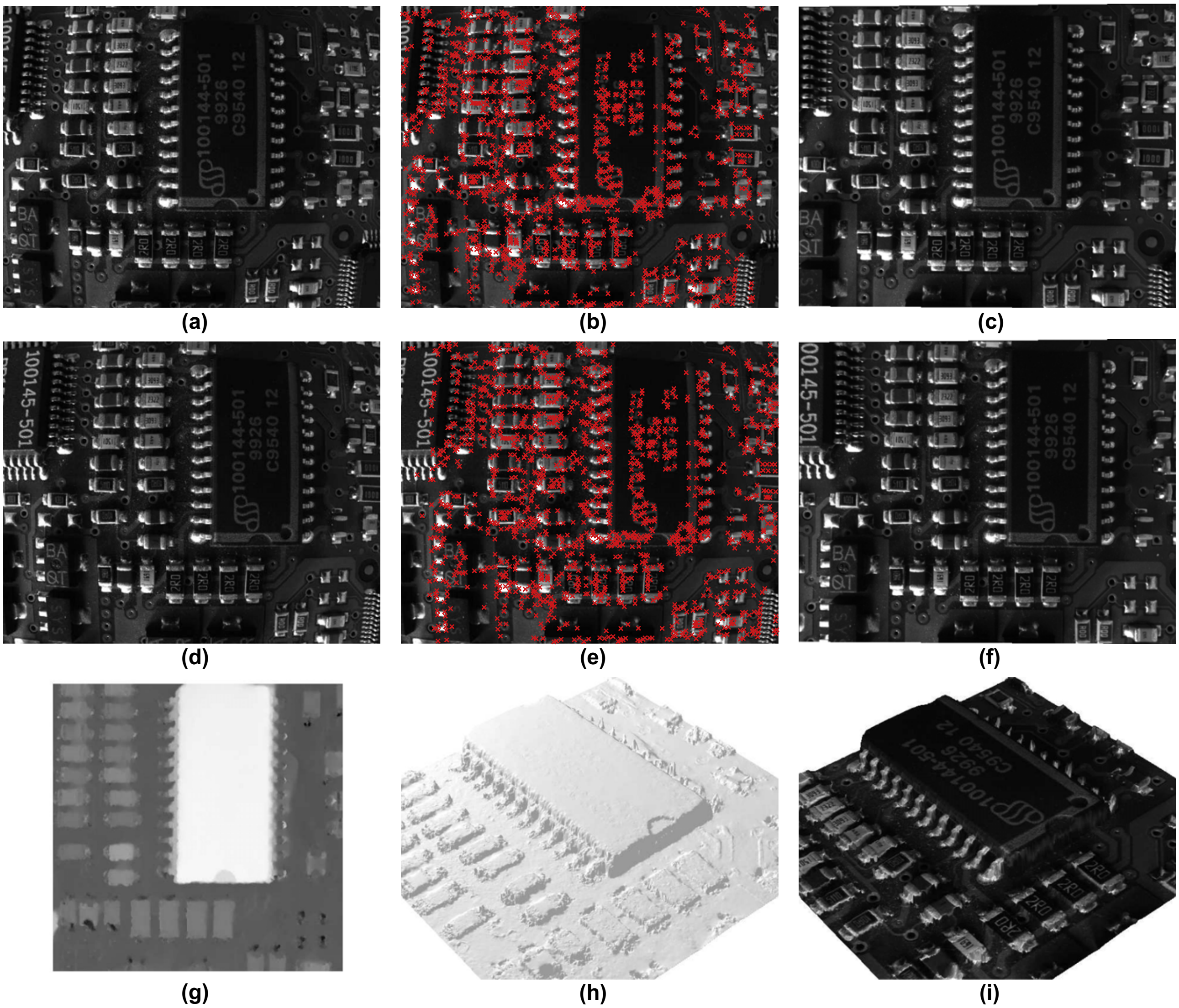


Fig. 6. Example for the automatic determination of the fundamental matrix and the radial distortion. (a) and (d) The input images. (b) and (e) The 1499 point matches determined by the algorithm in Section 8. (c) and (f) The images after undistortion and projective rectification. (g) The disparity computed from (c) and (f). (h) The disparity in (g) displayed as a 3D plot. (i) The disparity in (g) displayed as a 3D plot with the image in (c) overlaid as a texture.

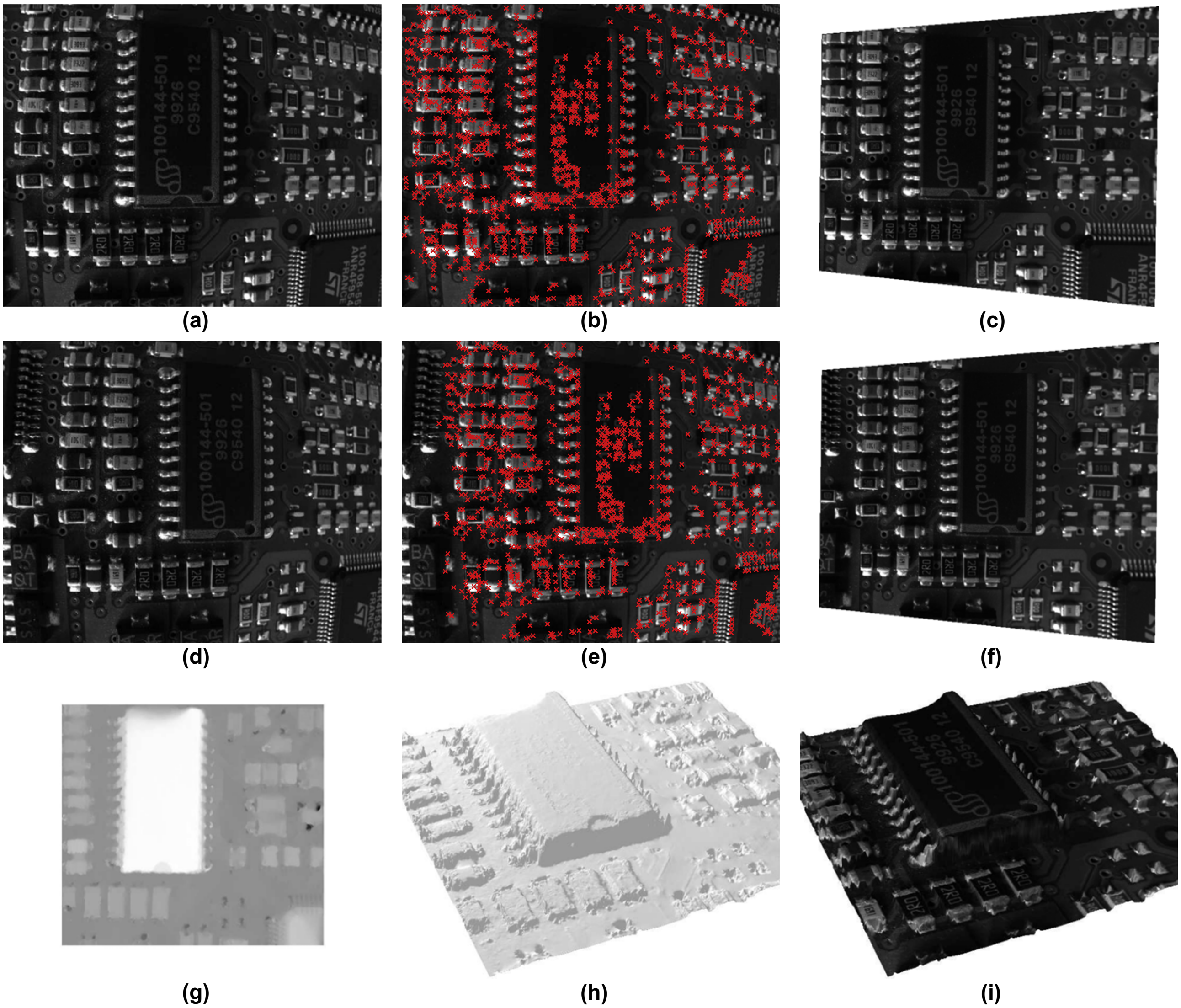
9.2.4. Ignoring distortions

To show the advantages of estimating the radial distortion in addition to the fundamental matrix, the same experiment as above was performed with an algorithm that determines the fundamental matrix under translation, but does not take into account radial distortions. The input images are shown in Figs. 8a and d (they are identical to Figs. 6a and c). The algorithm determined 583 point matches automatically with $t = 1$, which are shown in Figs. 8b and e. Figs. 8c and f display the corresponding projectively rectified images. Obviously, the radial distortion is still present in these images. The disparities computed from the rectified images are shown in Fig. 8g. Figs. 8h and i display them as 3D plots. Note that the radial distortions in the projectively rectified images lead to a severe bending of the reconstruction and to a significant number of outliers in the reconstruction. Overall, even the local shape of the components is not reconstructed well. Thus, modeling the radial distortion is essential to obtain correct reconstruction results in this example.

9.2.5. Comparison to solvers for general camera motions

Finally, we examine the question whether using an approach based on algorithms for the case of general camera motion is a viable alternative for determining the fundamental matrix under pure camera translations. To do so, we have implemented two different minimal solvers for the general motion case: the 8-point minimal solver proposed by Kúkelová et al. (2008) and the 9-point solver proposed by Fitzgibbon (2001). Furthermore, we have implemented the overdetermined linear solver proposed by Fitzgibbon (2001) as well as a Gold Standard solver. These algorithms were integrated into a RANSAC algorithm that is almost identical to the algorithm described in Section 8. The only modifications are to call the appropriate minimal solver and the appropriate algorithm for the overdetermined case.

To test the performance of the algorithms, six images of the PCB shown in Fig. 6 were acquired. In these images, the PCB was shifted from approximately 100 (between images 1 and 2) to 600 pixels (between images 1 and 6). The images in Figs. 6a and d are two



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 7. Example for the automatic determination of the fundamental matrix and the radial distortion with a tilted camera. (a) and (d) The input images. (b) and (e) The 1241 point matches determined by the algorithm in Section 8. (c) and (f) The images after undistortion and projective rectification. (g) The disparity computed from (c) and (f). (h) The disparity in (g) displayed as a 3D plot. (i) The disparity in (g) displayed as a 3D plot with the image in (c) overlaid as a texture.

images of this sequence. The three RANSAC algorithms (pure camera translation using the 3-point minimal solver and general camera translation using the 8-point and 9-point minimal solvers) were then applied to all 30 distinct image pairs (in our algorithm, matching image *i* to image *j* does not necessarily result in the same putative point correspondences as matching image *j* to image *i*.) The results of this experiment are displayed in Table 1.

As a first performance measure, we determined the average runtime of the initial RANSAC matching. This is a fair comparison since each algorithm is run with the same putative point correspondences. The runtime of the entire algorithm would not be a fair comparison because the runtime of the subsequent stages of the algorithm in Section 8 (guided matching and executing the RANSAC algorithm on the results of the guided matching) depends heavily on the quality of the results obtained by the initial RANSAC matching. Table 1 shows that the runtime of the general motion algorithms is approximately two orders of magnitude larger than the runtime of the pure translation algorithm. This comes as no surprise since the runtime of the algorithm is exponential in the number of point correspondences that must be used in the minimal solver. What is surprising, therefore, is that the runtime of the 9-point algorithm is better than the runtime of the 8-point algorithm. This can be explained by the fact that the 8-point algorithm finds significantly more viable solutions (solutions that fulfill (4) and (6)) per call than the 9-point algorithm, as can be seen from Table 1.

As an additional performance measure, we computed the mean value and standard deviation of the values of λ returned by the three different algorithms. Table 1 displays these statistics if the respective overdetermined linear solvers or the Gold Standard solvers are used to determine the fundamental matrix and radial distortion coefficient. Both the linear overdetermined and Gold Standard solvers for the pure translation case return very consistent results for all image pairs, as can be seen from the very small standard deviations of λ. In contrast, the algorithms for the general camera motion case return very inconsistent results, for which the standard deviation of λ has the same order of magnitude as λ itself. This is caused by the fact that the point matches in the images are highly ambiguous because of the large number of repeating structures in the images (in particular, the resistors, capacitors, and leads of the ICs). Apparently, the likelihood that the RANSAC algorithms for the general camera motion case return erroneous point correspondences is quite significant. This is not surprising since the algorithms for the case of general camera motions cannot constrain

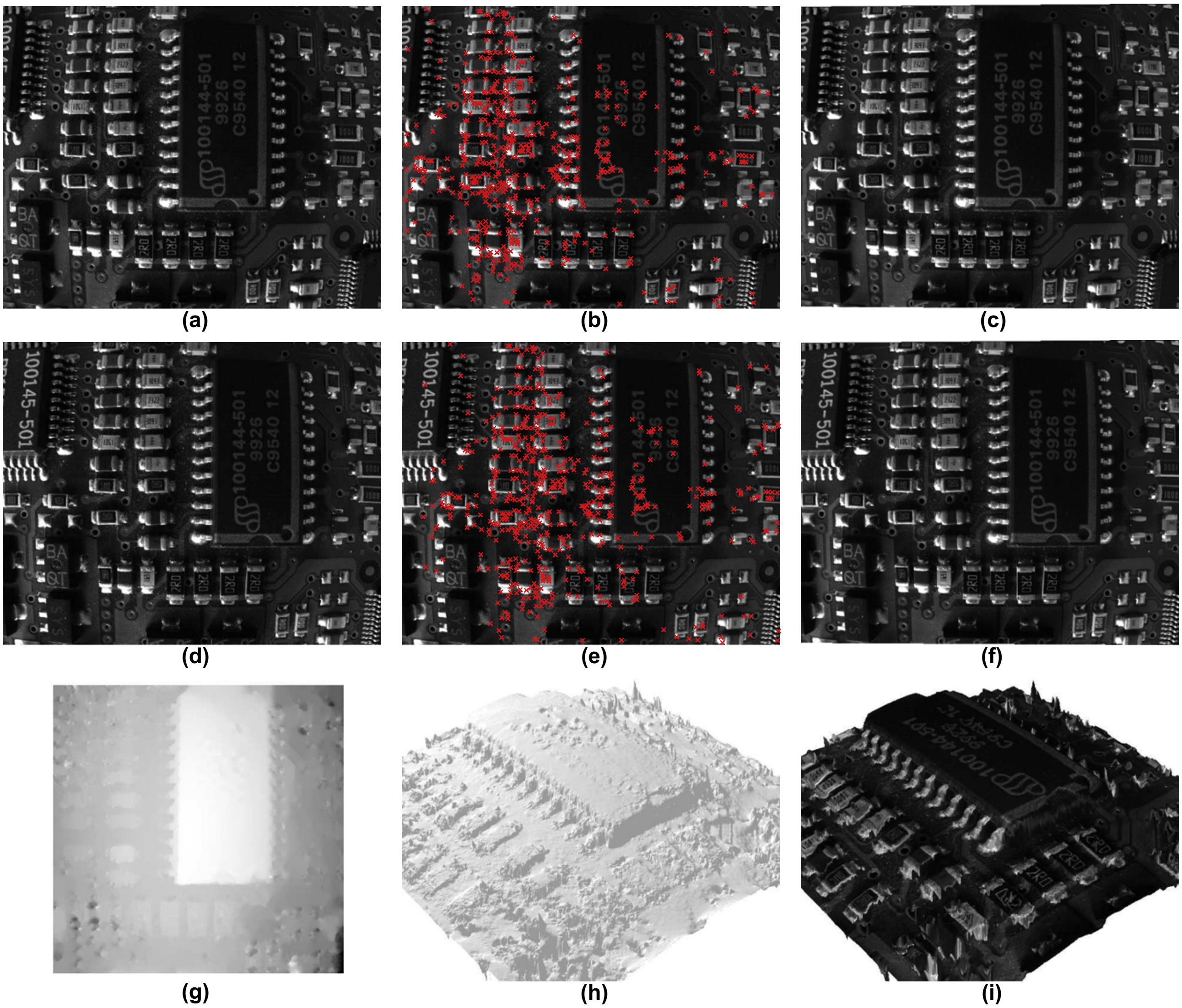


Fig. 8. Example for the automatic determination of the fundamental matrix without the determination of the radial distortion. (a) and (d) The input images. (b) and (e) The 583 point matches determined by a RANSAC algorithm that determines only the fundamental matrix under translation and not the radial distortion. (c) and (f) The images after projective rectification. (g) The disparity computed from (c) and (f). (h) The disparity in (g) displayed as a 3D plot. (i) The disparity in (g) displayed as a 3D plot with the image in (c) overlaid as a texture.

Table 1
Comparison of the results obtained with the pure translation solver (3-point) and two different general motion solvers (8-point and 9-point) on 30 image pairs obtained from a sequence of six images of a PCB. See text for details.

Minimal solver	3-Point	8-Point	9-Point
Average runtime for the initial RANSAC matching (ms)	30.9	3392.9	2699.4
Average number of solutions returned by the minimal solver	0.742	2.411	1.170
Mean value of λ (linear overdetermined solver) (1/pixel^2)	-2.133×10^{-7}	-3.464×10^{-7}	-4.102×10^{-7}
Standard deviation of λ (linear overdetermined solver) (1/pixel^2)	4.469×10^{-9}	1.686×10^{-7}	3.096×10^{-7}
Mean value of λ (Gold Standard solver) (1/pixel^2)	-2.098×10^{-7}	-2.694×10^{-7}	-2.959×10^{-7}
Standard deviation of λ (Gold Standard solver) (1/pixel^2)	2.891×10^{-9}	1.200×10^{-7}	1.725×10^{-7}

the matching geometry as well as the algorithm for the pure translation case. Therefore, we can see that using an algorithm that takes into account the correct imaging geometry is essential for obtaining correct results.

10. Conclusions

We have presented four algorithms for the estimation of the fundamental matrix $\mathtt{F}$ and the distortion coefficient λ under the division model for radial distortion for the case that the images are related by a pure camera translation.

We first derived a minimal three-point solver for the problem, which is essential to drive the RANSAC algorithm that determines $\mathtt{F}$ and λ automatically. To filter out incorrect solutions, we derived well founded practical bounds within which λ must lie.

We then extended the minimal solver to an overdetermined linear solver, which is essential to compute good approximate values for $\mathtt{F}$ and λ that can be used as starting values for the Gold standard solver. We also discussed potential instabilities of the overdetermined linear solver for very small values of λ and described solutions to prevent these instabilities.

In addition, we described a Gold Standard solver that returns the maximum likelihood estimate for $\mathbb{F}$ and λ. The Gold Standard solver should be used whenever results that are as accurate as possible are required.

Next, we showed that a reconstruction up to a similarity transform plus a scaling in the viewing direction is possible. We also described some structure degeneracies under which it is impossible to obtain a unique solution.

Finally, we combined the minimal solver with either of the two overdetermined solvers in a RANSAC algorithm determining $\mathbb{F}$ and λ automatically from interest points extracted from an image pair. We noted that it is essential to measure residual errors in the original, distorted image coordinate system.

We then performed experiments that show that the Gold Standard solver returns unbiased results with high accuracy and precision and that the overdetermined linear solver returns results that are biased for small radial distortions and large noise. We also established that the overdetermined linear solver nevertheless returns quite accurate results for realistic noise levels. In addition, we showed that the automatic determination of $\mathbb{F}$ and λ returns highly accurate 3D reconstructions (expressed by image disparities) in a real industrial application.

Finally, we compared the performance of the algorithms we have proposed for pure camera translations to algorithms for general camera motions. The results of these experiments show that using general camera motion algorithms for the pure translation case is not a viable alternative because the runtime of the general camera motion algorithms is two orders of magnitude larger than the runtime for the pure translation case. Furthermore, the robustness of the general camera motion algorithms to repetitive image structures is significantly lower.

The algorithms described in this paper are very efficient. The overdetermined linear solver runs in 0.52 ms for the 1499 point correspondences that the automatic solver has determined in the first example of Section 9.2. The Gold Standard solver requires 23.17 ms for the same data set. The automatic solver of Section 8, including matching of the point correspondences in the images, the RANSAC algorithm, and the guided matching, for this example (with 2349 and 2369 interest points in the two 1280 × 1024 images) requires 106.61 ms if the overdetermined linear solver is used and 234.17 ms if the Gold Standard solver is used. All times were measured on an Intel Core i5 750 CPU with 2.67 GHz.

Acknowledgements

I am grateful to Reviewer #2, whose comments helped me to discover the structure degeneracy for translations along the optical axis that is described in Section 7.

References

Armstrong, M., Zisserman, A., Beardsley, P., 1994. Euclidean structure from uncalibrated images. In: British Machine Vision Conference. pp. 509–518.

Brown, M., Hartley, R.I., Nistér, D., 2007. Minimal solutions for panoramic stitching. In: IEEE Conference on Computer Vision and Pattern Recognition. 8p (on CD-ROM).

Byröd, M., Kúkelová, Z., Josephson, K., Pajdla, T., Åström, K., 2008. Fast and robust numerical solutions to minimal problems for cameras with radial distortion. In: IEEE Conference on Computer Vision and Pattern Recognition. 8p (on CD-ROM).

Byröd, M., Brown, M., Åström, K., 2009. Minimal solutions for panoramic stitching with radial distortion. In: British Machine Vision Conference. pp. 41.1–41.11.

Drost, B., 2008. Ein variationsbasiertes Verfahren zum Stereosehen. Diplomarbeit, Fakultät für Mathematik, Technische Universität München.

Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24 (6), 381–395.

Fitzgibbon, A.W., 2001. Simultaneous linear estimation of multiple view geometry and lens distortion. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. I, pp. 125–132.

Förstner, W., 1994. A framework for low level feature extraction. In: Eklundh, J.-O. (Ed.), Third European Conference on Computer Vision, Lecture Notes in Computer Science, vol. 801. Springer-Verlag, Berlin, pp. 383–394.

Gluckman, J., Nayar, S.K., 2001. Rectifying transformations that minimize resampling effects. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. I, pp. 111–117.

Harris, C., Stephens, M., 1988. A combined corner and edge detector. In: Proceedings of The Fourth Alvey Vision Conference. pp. 147–151.

Hartley, R.I., Sturm, P., 1997. Triangulation. Computer Vision and Image Understanding 68 (2), 146–157.

Hartley, R., Zisserman, A., 2003. Multiple View Geometry in Computer Vision, second ed. Cambridge University Press, Cambridge.

Hornberg, A. (Ed.), 2006. Handbook of Machine Vision. Wiley-VCH, Weinheim.

Kúkelová, Z., Pajdla, T., 2007. A minimal solution to the autocalibration of radial distortion. In: IEEE Conference on Computer Vision and Pattern Recognition. 7p (on CD-ROM).

Kúkelová, Z., Bujnak, M., Pajdla, T., 2008. Automatic generator of minimal problem solvers. In: Forsyth, D., Torr, P., Zisserman, A. (Eds.), Tenth European Conference on Computer Vision, Lecture Notes in Computer Science, vol. 5304. Springer-Verlag, Berlin, pp. 302–315.

Kúkelová, Z., Byröd, M., Josephson, K., Pajdla, T., Åström, K., 2010. Fast and robust numerical solutions to minimal problems for cameras with radial distortion. Computer Vision and Image Understanding 114 (2), 234–244.

Lanser, S., Zierl, C., 1995. Robuste Kalibrierung von CCD-Sensoren für autonome, mobile Systeme. In: Dillmann, R., Rembold, U., Lüth, T. (Eds.), Autonome Mobile Systeme. Informatik aktuell. Springer-Verlag, Berlin, pp. 172–181.

Lanser, S., Zierl, C., Beutlhauser, R., 1995. Multibildkalibrierung einer CCD-Kamera. In: Sagerer, G., Posch, S., Kummert, F. (Eds.), Mustererkennung. Informatik aktuell. Springer-Verlag, Berlin, pp. 481–491.

Lenz, R., 1987. Linsenfehlerkorrigierte Eichung von Halbleiterkameras mit Standardobjektiven für hochgenaue 3D-Messungen in Echtzeit. In: Paulus, E. (Ed.), Mustererkennung. Springer-Verlag, Berlin, pp. 212–216.

Lenz, R., 1988. Viedeometrie mit CCD-Sensoren und ihre Anwendung in der Robotik. Habilitationsschrift, Lehrstuhl für Nachrichtentechnik der Technischen Universität München.

Lenz, R., Fritsch, D., 1990. Accuracy of videometry with CCD sensors. ISPRS Journal of Photogrammetry and Remote Sensing 45 (2), 90–110.

Li, H., Hartley, R., 2005. A non-iterative method for correcting lens distortion from nine point correspondences. In: Sixth Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras. 8p (on CD-ROM).

Li, H., Hartley, R., 2006. Five-point motion estimation made easy. In: 18th International Conference on Pattern Recognition, vol. 1, pp. 630–633.

Moons, T., Gool, L.V., Diest, M.V., Pauwels, E., 1993. Affine reconstruction from perspective image pairs obtained by a translating camera. In: Mundy, J.L., Zisserman, A., Forsyth, D. (Eds.), Applications of Invariance in Computer Vision, Lecture Notes in Computer Science, vol. 825. Springer-Verlag, Berlin, pp. 297–316.

Nistér, D., 2004. An efficient solution to the five-point relative pose problem. IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (6), 756–770.

Raguram, R., Frahm, J.-M., Pollefeys, M., 2009. Exploiting uncertainty in random sample consensus. In: 12th International Conference on Computer Vision. pp. 2074–2081.

Scherer-Negenborn, N., Schaefer, R., 2010. Model fitting with sufficient random sample coverage. International Journal of Computer Vision 89 (1), 120–128.

Steger, C., Ulrich, M., Wiedemann, C., 2008. Machine Vision Algorithms and Applications. Wiley-VCH, Weinheim.

Stewénius, H., 2005. Gröbner Basis Methods for Minimal Problems in Computer Vision. PhD Thesis, Centre for Mathematical Sciences, Lund University.

Stewénius, H., Engels, C., Nistér, D., 2006. Recent developments on direct relative orientation. ISPRS Journal of Photogrammetry and Remote Sensing 60 (4), 284–294.

Strand, R., Hayman, E., 2005. Correcting radial distortion by circle fitting. In: British Machine Vision Conference. pp. 9.1–9.10.