

Ioan Cleju · Dietmar Saupe

# Evaluation of Texture Registration by Epipolar Geometry

**Abstract** In the process of digitizing the geometry and appearance of 3D objects, texture registration is a necessary step that solves the 2D-3D mapping between the 2D texture images and the 3D geometric model. For evaluation of texture registration with ground truth, accurate datasets can be obtained with a complex setup consisting of calibrated geometry and texture capture devices. We do not have any knowledge of such evaluation performed; current evaluations reflect, at their best, the precision achieved by the algorithms, but fail to identify a possible bias. We propose a new evaluation measure based on the epipolar geometry of texture image pairs, with the advantage that the ground truth can be extracted solely from the texture images, independent of the 3D acquisition. We developed a noise model suitable to our purpose and analysed three distance measures based on epipolar geometry, well known in the computer vision community, to find new theoretical and experimental results. Finally, using the proposed framework, we evaluated a texture registration algorithm based on mutual information, and found that its accuracy was under half-pixel.

**Keywords** texture registration · epipolar geometry · epipolar distances · experimental evaluation · mutual information

---

Work supported by the DFG Research Training Group 1042 *Explorative Analysis and Visualization of Large Information Spaces* at the University of Konstanz.

Ioan Cleju  
Yotta Limited, Yotta House, 8 Hamilton Terrace, Leamington Spa, CV32 4LY, UK  
E-mail: ioan.cleju@yotta.tv

Dietmar Saupe  
Department of Computer and Information Science, University of Konstanz, Germany  
E-mail: dietmar.saupe@uni-konstanz.de

---

## 1 Introduction

As 3D scanning techniques become affordable, 3D reconstruction projects become widespread and numerous. Besides the geometric 3D shape, digitization may also include the surface appearance. Although many 3D acquisition devices are capable of capturing 3D geometry jointly with the surface appearance, the latter is more commonly generated from additional images of the model. The cameras used for 3D scanning are not suitable for high quality color images, and often the light conditions required for geometry acquisition are different than for texture acquisition. Instead, the surface appearance is captured in distinct colour images which are registered and fused onto a single texture map, suitable for real-time rendering on modern graphic cards. Texture registration is the process of aligning each image with the 3D model. Various algorithms have been successful for texture registration, but the field lacks a unified evaluation framework.

The goal of texture registration is to solve the 2D-3D mapping problem between each texture image and the 3D model. The mapping is most commonly modelled by a pinhole camera; the distortion-free pinhole camera is defined by the  $3 \times 4$ , rank 3 camera matrix  $P$ , and the fundamental equation  $x \sim PX$ , where  $\sim$  denotes equality in homogeneous coordinates, and  $X$ , respectively  $x$ , are the homogeneous coordinates of an arbitrary 3D point and its 2D projection.

The most common solution for texture registration uses 2D-3D feature point correspondences, often selected interactively; however, highly automatic algorithms based on a variety of criteria have been successfully proposed. The criteria include: point correspondences combined with outlines and color information [12], silhouettes [10], photo-consistency [3], mutual information [5], and even shadows [15]. The problem is formulated as an optimization for the camera parameters associated with each image and the algorithms typically are iterative in nature.

Evaluation of texture registration with ground truth requires highly accurate estimates for the camera matri-

ces associated with each image. Since the camera matrices are useful only if they are known in the reference system of the 3D model, the ground truth can be estimated when the texture acquisition cameras are calibrated with respect to the 3D scanning devices, requiring elaborate setup and imposing limitations on test datasets. Any estimation of the camera matrices using only the digitized 3D model and the texture images requires solving a 2D-3D registration problem, therefore is not suitable to our purpose.

Gold standard datasets for evaluation of 2D-3D medical image registration, with applicability restricted to this field, were produced using a highly complex, state-of-the-art medical imaging system [16]. The Stanford robot, with an arm that can be positioned with accuracy of 0.01 degrees, was used to build a test set for evaluation of multi-view stereo 3D reconstruction algorithms [13]. The high complexity and increased cost involved in producing accurate gold standard datasets motivated us to search for alternatives to ground truth estimation.

In this paper we propose a new measure for evaluation of texture registration, based on epipolar geometry, which overcomes the above-mentioned limitations. The advantage of our technique consists in the fact that the ground truth can be accurately obtained only from texture images, decoupling 3D scanning from texture acquisition, by full or weak calibration algorithms using known 3D patterns. The trade-off is that, rather than the full camera matrices, the method recovers a limited ground truth consisting in the epipolar geometry between images. The proposed error measure is not single-image based, such as the reprojection error (a common error measure for texture registration which will be defined later), but based on pairs of registered texture images. We will not discuss the extraction of the ground truth in the form of epipolar geometry, for which the literature is comprehensive, but define an appropriate error measure for texture registration using the limited ground truth.

We studied several epipolar distances and their quantitative relation to the reprojection distance. The distances are widely known in computer vision community, but we used them in a novel way. We analysed the epipolar distances under a statistical model different than the ones commonly found in literature, suitable to model the registration errors of texture registration algorithms. We found new theoretical and experimental aspects of epipolar distances, and confirmed others.

The proposed evaluation method is not limited to a particular algorithm; we evaluated a recent registration method using mutual information because the application, developed by the authors, was easily available. The algorithm proved to be highly accurate, with errors less than half-pixel.

In the remainder of the Introduction we will review the current methods and measures used for evaluation of 2D-3D registration, present our evaluation method, and, finally, we will give an overview of the rest for the paper.

### 1.1 Evaluation Measures for 2D-3D Registration

Texture registration is commonly evaluated by experiments consisting in repeated runs with random initialization. The *accuracy* characterizes the degree to which, by increasing the number of runs, the mean registration parameters converge to the correct parameters (ground truth). The *precision* characterizes the degree to which several registration runs give similar results. The accuracy can, therefore, be associated with the distance between the mean of the achieved parameters and the correct parameters, and the precision with the covariance matrix modelling the distribution of the achieved parameters.

It is convenient to characterize the quality of the registration by a single number, representing a geometric distance. This approach has some advantages, such as the fact that it is independent of the class of registration functions used (e.g., different camera models), and can be related geometric characteristics, for instance the image resolution. We mention two geometric distances, computed respectively in 2D image space and in 3D model space: the reprojection distance and the back projection distance. The reprojection distance (simply denoted 2D distance in some works) is a measure of the distances between the projections of the points of the model on the image plane with correct parameters, and those with the achieved parameters. As common, we will use the root mean squared distance (RMS). Thus, given two cameras  $P^1$  and  $P^2$  and a model consisting of  $n$  points  $\{X_i\}$ ,  $i = 1, \dots, n$ , their reprojection distance is:

$$D_{\text{rep}}(P^1, P^2, \{X_i\}) = \left( \frac{1}{n} \sum_{i=1}^n d(P^1 X_i, P^2 X_i)^2 \right)^{1/2} \quad (1)$$

The back projection distance is measured in 3D model space. Each sample point  $X_i$ ,  $i = 1, \dots, n$  is projected onto the image plane using the true camera projection parameters. The obtained image point is backprojected into the 3D scene, however, by using the approximate projection parameters. Generally, the resulting back projection ray misses the corresponding original sample point  $X_i$  and its distance from the ray is computed. Finally, the RMS distance of all sample points to their corresponding back projection rays yields the back projection distance. In literature, this distance can be found under different names as well, for instance projection distance [3] or reprojection distance [16].

The reprojection distance is the preferred error measure for texture registration [10, 5, 15], while the back projection distance is preferred for 3D reconstruction [7]; in this work, we will apply and study the reprojection error.

The accuracy and the precision can be defined for a geometric distance, as well. If the ground truth is known, we define the accuracy over an experiment as the RMS

distance from the projections with ground truth parameters to the projections with the mean parameters achieved in the experiment. We define the precision as the average of the RMS distances from the projections with the mean parameters of the experiment to the projections with the achieved parameters of the experiment.

We emphasize here that the accuracy is defined by the ground truth and cannot be estimated without it. The precision, on the other hand, is solely dependent on the experimental results of the algorithm, but cannot detect a consistent bias of the algorithm. High precision by itself does not insure high quality results, as the algorithm may suffer from the overfitting problem.

## 1.2 Proposed Evaluation

Camera matrices are reliably estimated by state-of-the-art calibration algorithms, using patterns with known Euclidean information [8]. When a calibration pattern is present in the scene, the camera calibration finds the camera parameters in the reference system of the pattern. A texture registration algorithm, on the other hand, finds the camera parameters in the reference system of the model. To compare the parameters obtained by camera calibration with those by texture registration, one needs the transformation between the pattern's and the model's reference systems. One way to achieve this is to pre-calibrate the 3D scanner (if possible) with respect to the pattern and obtain the model in the pattern's coordinate system; the same scene (object and pattern) would be used for 3D scanning and texture acquisition by photography.

In this article we propose an alternative method which allows comparing the approximate parameters of texture registration with those of a camera calibration algorithm *without* knowing the relation between the pattern's and the model's reference systems. The approach is flexible and completely decouples 3D scanning from texture acquisition and evaluation of registration. Our solution is a geometric distance based on *epipolar geometry*. The epipolar geometry of two images, characterized by the *fundamental matrix*, is invariant with respect to the scene.

The evaluation setup consists in a calibration pattern and the object of interest, kept fixed while the texture photographs are taken. On one hand, the texture images are registered to the model ignoring the pattern. On the other hand, each image is calibrated with respect to the pattern by camera calibration. For each image we obtain the camera matrix twice, once using camera calibration, in the reference system of the pattern, and using texture registration, in the reference system of the model. Consequently, for each image pair, the epipolar geometry can be estimated twice. We will call the fundamental matrix extracted using the pattern the *gold standard* and compare it with the data obtained from the registration algorithm. Since the epipolar geometry is independent of

the scene, the two fundamental matrices obtained for a given image-pair can be directly compared.

Alternatively, the fundamental matrices can be obtained by weak calibration, from a set of point-correspondences identified in both images, without requiring a pattern (though, the 3D locations of the points should not be in a degenerate configuration, such as a plane). The advantage of not having the pattern in the scene would be that the silhouette of the model on the image space could be larger. However, we chose to obtain the epipolar geometry from full calibration using Zhang's algorithm [19], implemented in the Bouguet toolbox [2] using a planar chessboard pattern, for the following reasons: firstly, the corner detection on a chessboard pattern is more accurate than detection of arbitrary features; secondly, using a pattern, the quality, number, and configuration of the feature points used in calibration are independent of the model's characteristics; lastly, the planar chessboard pattern is easy to construct.

The main focus of this article is finding a distance measure between the gold standard fundamental matrix and the two camera matrices obtained from texture registration, that approximates the reprojeciton distance<sup>1</sup> from (1) as accurate as possible. We re-used three distances widely known in computer vision community, namely the symmetric epipolar distance, the Sampson epipolar distance, and the manifold projection distance (a constrained variant of the total geometric distance to the joint image manifold), and evaluated them in a new statistical model. We found a theoretical bias of the manifold projection distance, closely confirmed by experiments, that is invariant with respect to scene configuration. Experiments showed that the biases of the other two distances depend on the scene configuration. We also found that the unbiased manifold projection distance estimates better the reprojeciton distance (1) than the other two distances.

## 1.3 Overview

In the following Section 2 we define the three epipolar distances and develop an algorithm to estimate the manifold projection distance for normalized cameras. In Section 3 we motivate the statistical model and present a theoretical result for the manifold projection distance, then develop an extensive experimental evaluation for the epipolar distances, in Section 4. In Section 5 we evaluate a texture registration algorithm based on mutual information using the proposed framework. Finally, we draw the conclusions in Section 6.

---

<sup>1</sup> We will extend this definition of the reprojeciton distance for two cameras.

## 2 Epipolar Geometry and Epipolar-based Distances

Before defining the epipolar distances, let us recall the basic notions of epipolar geometry. Let the images  $I$  and  $I'$  be obtained by two cameras with matrices  $P$  and  $P'$ , having distinct optical centers. If  $x$  and  $x'$  are the projections of the same 3D point  $X$  on  $I$  and  $I'$ , written in homogeneous coordinates, then the epipolar constraint

$$x'^T F x = x^T F^T x' = 0 \quad (2)$$

holds, where  $F$ , of dimension  $3 \times 3$ , is the fundamental matrix corresponding to the two images.

The *epipoles*,  $e$  in  $I$  and  $e'$  in  $I'$ , are the images of the camera centers  $C'$  and  $C$ , respectively, on the other camera. The product  $Fx$  defines an *epipolar line* on image  $I'$  and  $F^T x'$  an epipolar line on  $I$ . All epipolar lines of  $I'$  are concurrent on the epipole  $e'$ . Similarly, the epipolar lines on  $I$  are concurrent on  $e$ . The constraint (2) states that  $x'$  should be on the epipolar line  $Fx$  and  $x$  should be on the epipolar line  $F^T x'$ .

The camera matrix  $P$  can be decomposed into the product of the  $3 \times 3$  *camera calibration matrix*  $K$  and the  $3 \times 4$  roto-translation matrix  $[R|T]$ . The matrix  $[R|T]$  is the projection matrix of the *normalized camera* corresponding to  $P$ . A normalized camera is a camera with  $K$  the identity matrix. The camera projection matrix  $[R|T]$  has fewer degrees of freedom (six) than the general matrix  $P$ , and a geometrical interpretation as a roto-translation.

The fundamental matrix of a pair of normalized cameras is called the *essential matrix*, denoted  $E$ . It is a specialization of the fundamental matrix, and the following relation holds:  $E = K'^T F K$ . Both the fundamental and essential matrices are homogeneous quantities.

The *joint image space* is the cartesian product of all point pairs in two images, and is a subset of  $\mathbb{R}^4$ . Given the fundamental matrix between the two images, the *joint image manifold* is the subset of the joint image space, composed of pairs matching according to the epipolar geometry (2). The joint image manifold is a three dimensional manifold embedded in  $\mathbb{R}^4$ . It is also an *algebraic manifold* of degree two, because it is the zero set of the second degree polynomial defined by (2).

### 2.1 Epipolar Distances

Let a 3D model be defined through a sampling  $\{X_i \mid i = 1, \dots, n\}$ . Let the camera matrices corresponding to the images  $I$  and  $I'$ , obtained from the texture registration algorithm, be  $P^1$  and  $P^{1'}$ , the fundamental and essential matrices be  $F^1$  and  $E^1$ , and the projections of  $X_i$  on  $I$  and  $I'$  be  $(x_i, x'_i)$ ,  $i = 1, \dots, n$ . Let the gold standard be the fundamental and essential matrices  $F^2$  and  $E^2$ . In this subsection we abbreviate for shortness  $F := F^2$ .

We will define several geometric distances between the gold standard and the configuration obtained by texture registration, all based on epipolar geometry. The same formalism stands for general cameras as well as for normalized cameras, in the latter case replacing the fundamental matrix with the essential matrix.

*Symmetric Epipolar Distance.* Considering the geometric meaning of the product  $Fx_i$ , the symmetric epipolar distance is defined by the root mean of the summed squared distances between each image point  $x'_i$  and the epipolar line  $Fx_i$ , and, respectively, between  $x_i$  and  $F^T x'_i$ :

$$\begin{aligned} D &= \left( \frac{1}{2n} \sum_{i=1}^n d(x'_i, Fx_i)^2 + d(x_i, F^T x'_i)^2 \right)^{1/2} \\ &= \left( \frac{1}{2n} \sum_{i=1}^n \frac{(x'^T F x_i)^2}{(Fx_i)_1^2 + (Fx_i)_2^2} + \frac{(x_i^T F^T x_i)^2}{(F^T x'_i)_1^2 + (F^T x'_i)_2^2} \right)^{1/2} \end{aligned} \quad (3)$$

*Sampson Epipolar Distance.* The equation defining the joint image manifold corresponding to the fundamental matrix  $F$  is  $A(x, x') = 0$ , with  $A(x, x') = x'^T F x$ ; for any pair  $(x, x') \in \mathbb{R}^4$ ,  $A$  can be linearized using the Jacobian  $J$  of  $A$  (a row vector of dimension 4). Denoting  $A_i = A(x_i, x'_i)$  and  $J_i = J(x_i, x'_i)$ ,  $i = 1, \dots, n$ , the Sampson distance (for the  $i$ -th sample point) is defined as the minimal norm of the vector  $\delta_{x_i, x'_i}$ , where  $\delta_{x_i, x'_i}$  satisfies  $J_i \delta_{x_i, x'_i} = -A_i$ . This is the condition for the first order Taylor expansion of  $A(x, x')$ , at  $(x_i, x'_i) + \delta_{x_i, x'_i}$ , to be 0. Thus, the Sampson epipolar distance is a first order approximation of the distance from  $(x_i, x'_i)$  to the joint image manifold. The total Sampson distance, for all projection pairs, is:

$$\begin{aligned} D &= \left( \frac{1}{2n} \sum_{i=1}^n \frac{A_i^2}{J_i J_i^T} \right)^{1/2} \\ &= \left( \frac{1}{2n} \sum_{i=1}^n \frac{(x'^T F x_i)^2}{(Fx_i)_1^2 + (Fx_i)_2^2 + (F^T x'_i)_1^2 + (F^T x'_i)_2^2} \right)^{1/2} \end{aligned} \quad (4)$$

*Manifold Projection Distance.* The third distance we discuss is a variant of the geometric distance (reprojection distance for fundamental matrix estimation [8]), constrained to cameras with given fundamental matrix, denoted here as the manifold projection distance. We define the manifold projection distance using the reprojection distance (1), as the minimum value of the following expression constrained by cameras  $P^2$  and  $P^{2'}$  that yield the gold standard fundamental matrix  $F^2$ :

$$\begin{aligned} D &= \min_{F(P^2, P^{2'})=F^2} \left( \left( \frac{1}{2} D_{rep}(P^1, P^2, \{X_i\})^2 \right. \right. \\ &\quad \left. \left. + \frac{1}{2} D_{rep}(P^{1'}, P^{2'}, \{X_i\})^2 \right)^{1/2} \right) \end{aligned} \quad (5)$$

As the projections obtained with the optimal camera matrices  $P^2$  and  $P^{2'}$  are on the joint image manifold defined by the fundamental matrix  $F^2$ , the total distance to the joint image manifold is a lower bound of the manifold projection distance.

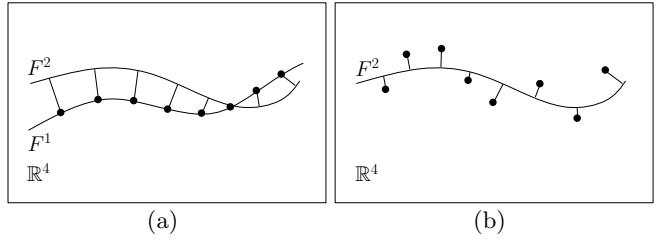
## 2.2 Discussion

We will relate our problem to the well known problem of fundamental matrix estimation from noisy point correspondences in two images [8]. Several epipolar distances have been proposed as objective functions for estimation of the fundamental matrix from point correspondences. We mention the algebraic distance, the symmetric epipolar distance, the Sampson distance (gradient-weighted epipolar distance), and the reprojection distance. The algebraic distance is defined as the sum of squared residuals  $x_i^T F x_i$  from (2). The symmetric epipolar distance and the Sampson distance were defined above in (3), respectively (4). The reprojection distance is defined by the same expression as the manifold projection distance (5), but the optimization space differs; the noisy coordinates of the projections,  $(x_i, x'_i)$ , are known, and the optimization is performed for the essential matrix and the structure  $X_i$ ,  $i = 1, \dots, n$ .

In our case, an estimated 2-camera description and the 3D structure are given, and the problem is to compare this configuration with a highly accurate essential matrix. The distances used as objective functions for estimation of the essential matrix can be used to define a suitable distance measure for our problem. The differences between the two problems consist, firstly, in the different known and unknown parameters, and secondly, in the statistical noise model. For estimation of fundamental matrix, the imprecision is in the extraction of 2D coordinates, and the 3D structure is unknown, while in our case the noise is associated with inaccurately estimated camera matrices, and the 3D structure is known.

The algebraic distance, the symmetric epipolar distance, and the Sampson epipolar distance are applied in exactly the same manner in both problem instances. The reprojection error requires the 3D structure. As for our instance of the problem the structure is known, the search space consists in the parameters which, together with an essential matrix, identify the twelve free parameters of a pair of normalized cameras. For essential matrix estimation using the reprojection error, the search space consists in the five free parameters of the essential matrix and the  $3n$  parameters defining  $n$  point coordinates in 3D space (the structure).

We illustrate the conceptual difference between the two problems in Figure 1. For our evaluation problem, the point correspondences lie on the joint image manifold with essential matrix  $F^1$ , and we estimate the distance to the manifold defined by the gold standard  $F^2$ , with the additional constraint that the solution should be consistent with the given 3D structure  $X_i$ ,  $i = 1, \dots, n$ . This is



**Fig. 1** Manifold projection distance (a). The point correspondences are on the manifold defined by  $F^2$ , and we estimate the total distance to the manifold defined by  $F^1$ . Fundamental matrix estimation from noisy point-correspondences (b).

the motivation behind the name manifold projection distance. For the fundamental matrix estimation problem, the point correspondences do not lie on a manifold, and the essential matrix  $F$  defining the 'closest' manifold is searched for.

An alternative solution to our problem would be to use a matrix distance between two fundamental matrices. It has been shown in literature that a matrix distance, such as the Frobenius norm of the difference of two (normalized) fundamental matrices, is not appropriate to measure the difference between fundamental matrices [17, 6]. Zhang gives an algorithm<sup>2</sup> to compute a geometric distance between two fundamental matrices [17]. We believe that, having the 3D structure and a completely defined camera pair known, this distance is absorbed by the symmetric epipolar distance.

Several studies [14, 17, 18] considered various epipolar distances for estimation of the fundamental matrix. The algebraic distance performed the worst [17, 8], as expected, since it does not reflect a geometric measure. The Sampson distance is considered to be better than the symmetric epipolar distance, as it is the first order approximation of the geometric distance, fact that was confirmed in experiments [17].

In [18] the symmetric epipolar distance, the Sampson distance, and the reprojection distance were compared with respect to their use as optimization criteria, when the coordinates of the point correspondences are affected by i.i.d. Gaussian noise. The symmetric epipolar distance and the Sampson distance were equivalent only when the epipoles are at infinity and the images have the same scale of the scene. On the other hand, the reprojection distance and the Sampson distance were equivalent when the epipoles are at infinity, and slightly different when the epipoles are in the images.

Other noise models have been considered for fundamental matrix estimation, such as heteroscedastic noise [9], or presence of outliers [14, 17]. For our problem, the geometry is known and the 2D projections are generated

<sup>2</sup> Initially proposed by Stéphane Laveau from INRIA Sophia-Antipolis, and mentioned also in [6].

by projecting the 3D structure, therefore we do not face any outlier problem. We associate the noise model with the estimation of camera matrix, which generates correlated noise in the 2D projections over images. Likewise, it is beyond the scope of this paper to discuss the estimation of the golden standard fundamental matrix.

### 2.3 Computation of the Manifold Projection Distance

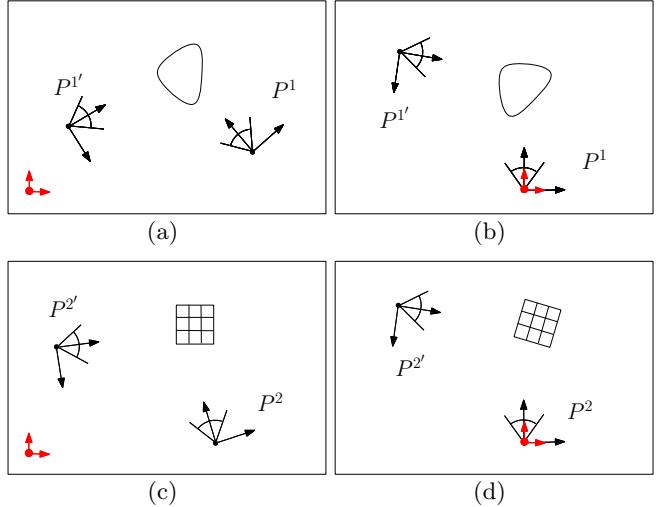
While the symmetric and Sampson epipolar distances are computed straightforward from equations (3) and (4), the manifold projection distance (5) requires a complex optimization on the subspace of two cameras with a given essential matrix. In this section we provide an algorithm for that.

*Transformation to Canonical Form of Camera Pairs.* Let the projection matrices of two arbitrary normalized cameras be  $P = [R|T]$  and  $P' = [R'|T']$ , where  $R$ ,  $R'$ ,  $T$ , and  $T'$  are rotation matrices and translation vectors, respectively (we exclude cameras with the same optical centre). Their *canonical form* is the camera pair  $([I|0], [R^c|T^c])$  with the same essential matrix  $E$  [8]. Then, the following homogeneous relation holds:

$$E \sim [T^c] \times R^c \quad (6)$$

The five degrees of freedom of the essential matrix can be pointed out in equation (6): the right-hand side of the equation has six degrees of freedom; however, due to the scale ambiguity of the essential matrix, the scale of the Euclidean space (the scale of  $T^c$ ) is 'lost'. Given an essential matrix, a corresponding canonical pair of cameras  $([I|0], [R^c|T^c])$  can be extracted, however, the result is not unique. Instead, there are four solutions for  $R^c$ ; as well, the scale of  $T^c$  is undetermined. In many practical situations it is easy to choose the desired solution for  $R^c$ : a point known to be visible in both images is 'in front' of both cameras in one configuration only. However, in our evaluation framework, we avoided this problem, as we directly transformed the camera-pair obtained by camera calibration into the canonical form; this was possible because we obtained the essential matrix indirectly, after full camera calibration.

As a first step we transform the given camera pair  $([R^1|T^1], [R'^1|T'^1])$  (obtained by texture registration) to the canonical form  $([I|0], [R^{1c}|T^{1c}])$ . To keep the consistency of the scene, we transform the 3D model correspondingly, multiplying all point coordinates with the  $4 \times 4$  matrix obtained from  $[R^1|T^1]$  padded with the vector  $(0, 0, 0, 1)$  as the last row. Projecting the new coordinates of the model using the canonical pair  $([I|0], [R^{1c}|T^{1c}])$  is equivalent to projecting the old coordinates using the original camera pair  $([R^1|T^1], [R'^1|T'^1])$ . In a second step we extract the proper canonical camera pair  $([I|0], [R^{2c}|T^{2c}])$  from the essential matrix  $E^2$  (choosing the appropriate solution out of the four possibilities). See Figure 2.



**Fig. 2** The normalized cameras pairs  $(P^1, P^{1'})$ , in the object's reference system (a), and  $(P^2, P^{2'})$ , in the pattern's reference system (c), are transformed to the canonical form (b), respectively (d). The object and the pattern are transformed to the reference system of the canonical form. The epipolar geometry, as well as the object's and pattern's projections, are not changed by this transformation.

If the texture registration and the essential matrix estimation would be exact, then  $R^{1c} = R^{2c}$  and  $T^{1c} = sT^{2c}$  with a non-zero scale factor  $s$  ( $s = 0$  would imply that the optical centers coincide).

*Rewriting the Minimization.* The search space of (5) is the set of all camera pairs with the same essential matrix. Since the space of two cameras has 12 degrees of freedom and the essential matrix 5, the search space has 7 degrees of freedom, and it is  $SE(3) \times \mathbb{R}^*$ . Indeed, one degree of freedom consists in the non-zero scale  $s$  which defines the Euclidean space scale (which was 'lost' in the essential matrix), and the other six correspond to the rotation and translation of the scene with respect to the canonical pair, written together in one transformation as

$$H_s^{op} = \begin{pmatrix} R^{op} & | & T^{op} \\ 0 & | & s \end{pmatrix} \quad (7)$$

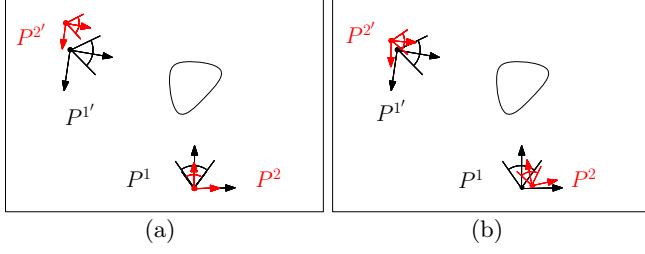
Thus, our optimization problem (5) can be rewritten as

$$D = \min_{H_s^{op}} \left( \frac{Q}{2n} \right)^{1/2} \quad (8)$$

where

$$\begin{aligned} Q = \sum_{i=1}^n d(x_i, [I|0]H_s^{op}X_i)^2 + d(x'_i, [R^{2c}|T^{2c}]H_s^{op}X_i)^2 \\ = \sum_{i=1}^n d(x_i, [I|0]H_1^{op}X_i)^2 \\ + d(x'_i, [R^{2c}|sT^{2c}]H_1^{op}X_i)^2 \end{aligned} \quad (9)$$

and the minimization is carried out over all transforms of the form (7).



**Fig. 3** The camera pair  $(P^1, P^{1'})$  has the essential matrix  $E^1$ , and the camera pair  $(P^2, P^{2'})$ ,  $E^2$ . The initialization (a) reduces to 0 the residuals of (8) corresponding to one image. The terms of (8) coming from both images are optimized (b). The essential matrices are unchanged.

*Iterative Optimization.* The optimization for  $H_s^{op}$  is carried out iteratively, requiring a sufficiently precise initialization. The equation (9) does not contain  $s$  in the terms corresponding to the first image. Initializing  $RT_1^{op}$  with  $R^1$  and  $T^1$  reduces the error from the first image to 0, and we use this initialization to easily find the value of  $s$ , minimizing the term corresponding to the second image,  $\sum_{i=1}^n d(x'_i, [R^{2c}|sT^{2c}]H_1^{op}X_i)^2$ , by solving a least squares problem. We solved it by iterated linear least squares. Two steps were sufficient.

The result serves as a good initialization for the global optimization over all seven degrees of freedom, allowing errors on both terms, by an iterative scheme; we used the Levenberg-Marquardt algorithm. The pseudo-code is sketched in Algorithm 1 and illustrated in Figure 3. The scale  $s = 0$  is not a singularity of (9), therefore the optimization is carried out safely on the manifold  $SE(3) \times \mathbb{R}$ .

During the iterative optimization, for any value of  $H_1^{op}$  we constructed a local parametrization of  $SE(3)$  on  $\mathbb{R}^6$  using the Lie algebra  $se(3)$  associated with  $SE(3)$ , and we performed the iteration in  $\mathbb{R}^6 \times \mathbb{R}$ . Compared to other parametrizations, such as Euler angles for the rotation matrix, this framework does not suffer from problems as gimbal lock. Similar optimization problems can be found in algorithms for motion and structure; for instance, optimization techniques for Riemannian manifolds, applied for optimization on the essential manifolds [11].

#### Algorithm 1 Estimation of the Manifold Projection Distance

---

```

Initialize  $x_i = [R^1|T^1]X_i$  and  $x'_i = [R^{1'}|T^{1'}]X_i$  for  $i = 1, \dots, n$ 
Extract camera pair  $[I|0]$  and  $[R^{2'}|T^{2'}]$  from  $E^2$ 
Initialize  $H_1^{op}$  with  $[R^1|T^1]$ 
Minimize  $Q$  in (9) for  $s$ , iterated linear least squares
Minimize  $Q$  in (9) for  $(H_1^{op}, s)$ , Levenberg-Marquardt
Return  $D$  (8)

```

---

### 3 Theoretical Analysis

Texture registration algorithms solve the camera matrix  $P$  in the reference system of the model, for each texture image. Typically, the camera calibration matrix  $K$  is found separately by camera calibration, therefore we consider that texture registration effectively estimates  $[R|T]$ . We will search for a meaningful statistical model of matrices  $[R|T]$  obtained by a generic, unbiased registration algorithm, then analyse the evaluation distances in this context.

Texture registration is solved as an optimization problem, and let  $\Psi([R|T], X, I)$  be a generic objective function. The unknown element is the matrix  $[R|T]$ , while the image  $I$  and 3D model  $X = \{X_i | i = 1, \dots, n\}$  are known. Let  $x$  be the  $2n$ -vector containing the projections of the points  $X_i$ ,  $i = 1, \dots, n$ , on the image space, concatenated. Since the model  $X$  is fixed, the number of free parameters of  $x$  is 6, corresponding to the degrees of freedom of  $[R|T]$ .

A typical texture registration algorithm evaluates  $\Psi$  in two steps: first, it projects the points to the 2D space using the camera projection, then evaluates  $\Psi(x, I)$ . The objective function depends indirectly on the projection matrix, but is effectively evaluated using the 2D projections. An illustrating example is registration based on 2D-3D point correspondences, where the objective function is the distance between projected 2D coordinates ( $x$ ) and 2D coordinates extracted from  $I$ . It is natural to consider that the distribution of the solutions found in the optimization of  $\Psi(x, I)$  over  $x$  is a Gaussian centered in the correct 2D coordinates  $\tilde{x}$ , isotropic in the 6-dimensional manifold space around  $\tilde{x}$ . Consequently, approximating the camera projection locally by an affine operator, the statistical model of  $[R|T]$  will be a Gaussian centered in the correct projection matrix.

It is easy to check that the total covariance of  $x_i$  (the covariance of  $x$ ) is related to the expected root mean squared error (RMSE) of the projections by the following relation:

$$E(RMSE(x_i, \tilde{x}_i))^2 = \frac{1}{n} \sum_{i=1}^n \text{tr}(\text{cov}(x_i)) \quad (10)$$

$$= \frac{1}{n} \text{tr}(\text{cov}(x)) \quad (11)$$

Our goal is to compare, theoretically and experimentally, the reprojection error (RMSE) and the epipolar distances. Since the former distance is defined for one camera and the latter for one camera-pair, we extend the reprojection distance (1) to a camera-pair, and refer to it as the *reference distance*:

$$D_{\text{ref}}((P^1, P^{1'}), (P^2, P^{2'}), \{X_i\}) = \quad (12)$$

$$\left( \frac{1}{2} D_{\text{rep}}(P^1, P^2, \{X_i\})^2 + \frac{1}{2} D_{\text{rep}}(P^{1'}, P^{2'}, \{X_i\})^2 \right)^{1/2}$$

Under the distribution considered above, we will prove a theoretical bias of the manifold projection distance with respect to the reference distance.

### 3.1 Expected Bias of the Manifold Projection Distance

For an arbitrary 3D model  $\{X_i | i = 1, \dots, n\}$ , we define the function  $f : \mathbb{R}^{24} \rightarrow \mathbb{R}^{4n}$ , parameterized by the camera pair  $(P, P')$ , to be the vector of the 2D projections of  $X_i$ ,  $i = 1, \dots, n$  on both image planes. If we restrict the domain of  $f$  to normalized cameras, the range of  $f$  is a 12-dimensional manifold  $S$  embedded in  $\mathbb{R}^{4n}$ . Spanning only the 7 degrees of freedom of two normalized cameras with fixed essential matrix  $E$ ,  $f$  will generate a 7-dimensional manifold  $S_E$  embedded in  $S$ .

Denoting  $Y_1 = f(P^1, P^{1'})$ ,  $Y_2 = f(P^2, P^{2'})$ , the reference distance (12) can be expressed using the Euclidean distance in  $\mathbb{R}^{4n}$ ,

$$D_{\text{ref}}((P^1, P^{1'}), (P^2, P^{2'}), \{X_i\}) = \left( \frac{1}{2n} \right)^{1/2} \|Y_1 - Y_2\|_2 \quad (13)$$

and the manifold projection distance can be rewritten as

$$D_3((P^1, P^{1'}), (P^2, P^{2'}), \{X_i\}) = \min_{\hat{Y} \in S_E} \left( \frac{1}{2n} \right)^{1/2} \|Y_1 - \hat{Y}\|_2 \quad (14)$$

where  $E$  is the essential matrix corresponding to  $(P^2, P^{2'})$ .

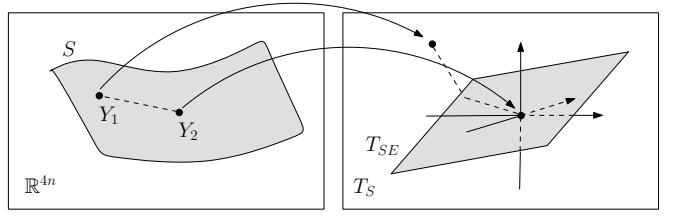
The optimal  $\hat{Y}$  is the orthogonal projection of  $Y_1$  on  $S_E$ .

Let  $T_S$  be the (12-dimensional) tangent space of  $S$  in  $Y_2$ , which is image of the affine approximation of  $f$  around  $(P^2, P^{2'})$ . The Gaussian distribution  $\mathcal{N}((P^2, P^{2'}), \Sigma)$  is mapped by the affine approximation of  $f$  into the Gaussian distribution  $\mathcal{N}(0, \Sigma_{Y_2}^S)$  in  $\mathbb{R}^{12}$ . The noise model for camera parameters presented in the previous section is equivalent to  $\mathcal{N}(0, \Sigma_{Y_2}^S)$  being isotropic, with total variance  $2\delta$ , where  $\delta$  is the total variance of projections in one image. Let  $T_{SE}$  be the (7-dimensional) subspace of  $T_S$  corresponding to cameras with essential matrix  $E$ .

Approximating  $f$  affine around  $(P^2, P^{2'})$  simplifies the analysis of (14), as the minimizer  $\hat{Y}$  is the Euclidean projection of  $Y_1$  on  $T_{SE}$  (Figure 4). Since the probability distribution used to generate  $Y_1$ , described in the previous section, is an isotropic Gaussian on the 12-dimensional space  $T_S$ , and  $T_{SE}$  is a 7-dimensional space, the relative bias of the manifold projection distance with respect to the reference distance is:

$$E \left[ \frac{\min_{\hat{Y} \in S_E} \|Y_1 - \hat{Y}\|_2}{\|Y_1 - Y_2\|_2} \right] \approx \left( \frac{12 - 7}{12} \right)^{1/2} \quad (15)$$

The result states that the total reprojection distance can be decomposed into a part that corresponds to the



**Fig. 4** Decomposition of the segment  $[Y_1, Y_2]$  in the tangent space  $T_S$ , according to the subspace  $T_{SE}$ .

5-dimensional manifold subspace associated with the degrees of freedom of the essential matrix, estimated by the manifold projection distance, and a distance that corresponds to the remaining 7-dimensional subspace. Under the statistical model we considered, the bias is a direct consequence of the number of degrees of freedom of the data associated with the distances. Another consequence is that no other distance based only on the essential matrix can estimate the reprojection distance better than the unbiased manifold projection distance.

For (15), we used the result that the projection of an  $n$ -dimensional isotropic Gaussian distribution of total variance  $n\sigma^2$  onto an  $s$ -dimensional subspace has total variance  $s\sigma^2$ . A similar proof is provided in [8] for estimation of the RMS residual error corresponding to Maximum Likelihood Estimators.

We expect, therefore, the manifold projection distance to have a bias around 0.645 with respect to the reference distance.

## 4 Experimental Analysis

We ran experiments to compare the proposed epipolar distances with the reference distance. Our evaluation framework was based on the following procedure, repeatedly evaluated in a variety of configurations:

1. Sample a 3D model into a set of points  $\{X_i | i = 1, \dots, n\}$ ;
2. Generate two normalized camera matrices:  $P^2 = [R^2 | T^2]$  and  $P^{2'} = [R^{2'} | T^{2'}]$ ;
3. Estimate a Gaussian noise model (covariance matrices) for  $P^2$  and  $P^{2'}$ ;
4. Generate a random pair  $P^1$  and  $P^{1'}$  according to the noise model;
5. Compute the reference distance  $D_{\text{ref}}$  between the two pairs according to (12);
6. Extract the essential matrix  $E^2 = E(P^2, P^{2'})$ , and compute the distances  $D((P^1, P^{1'}), E^2)$  using the epipolar distances; let  $D_1$ ,  $D_2$ ,  $D_3$ , be the symmetric epipolar distance, the Sampson distance, and the manifold projection distance, respectively;
7. Compare the distances  $D_1$ ,  $D_2$ , and  $D_3$  with  $D_{\text{ref}}$ .

The primary indicators we analysed were the ratios  $R_k = \frac{D_k}{D_{\text{ref}}}$ , their means  $\bar{R}_k$  and normalized standard deviations  $\sigma_{rk} = \frac{\sigma_{Rk}}{\bar{R}_k}$ ,  $k = 1, 2, 3$ . The means represent the relative biases of the epipolar distances with respect to the reference distance, and the normalized standard deviations give the approximation power of the unbiased epipolar distance with respect to the reference distance.

#### 4.1 Noise Model

We developed a numerical method to obtain a diagonal covariance matrix of camera parameters<sup>3</sup> according to Section 3, given a total variance of the projections vector:

$$\sum_{i=1}^n \text{tr}(\text{cov}(x_i)) = \delta \quad (16)$$

The expected RMSE of the reprojection error is  $\delta/n$ .

Let a normalized camera have the projection matrix  $[R|T]$ , and a model  $\{X_i | i = 1, \dots, n\}$ . Let the Gaussian random variable  $P = [R_p|T_p]$  have the distribution  $\mathcal{N}([R|T], \Sigma_{[R|T]})$ . For any projection  $x_i = PX_i$ , where  $P$  is the projection matrix of a camera and  $X_i$  a 3D point, we linearize the projection around  $\bar{P} = [R|T]$ , using the Jacobian of  $x_i$ ,  $J_{Pi}$ , with respect to  $P$ , and obtain a relation between the covariance of  $x_i$  and that of  $P$ :

$$\begin{aligned} \text{cov}(x_i) &= E[(x_i - \bar{x}_i)(x_i - \bar{x}_i)^T] \\ &\approx E[J_{Pi}(P - \bar{P})(P - \bar{P})^T J_{Pi}^T] \\ &\approx J_{Pi} \text{cov}(P) J_{Pi}^T \end{aligned} \quad (17)$$

For simplicity, we use the same notation  $P$  for the random variable and for a sample of it. Moreover, in (17) we denote by  $P - \bar{P}$  the appropriate column vector corresponding to the ordering of variables in the definition of the Jacobian.

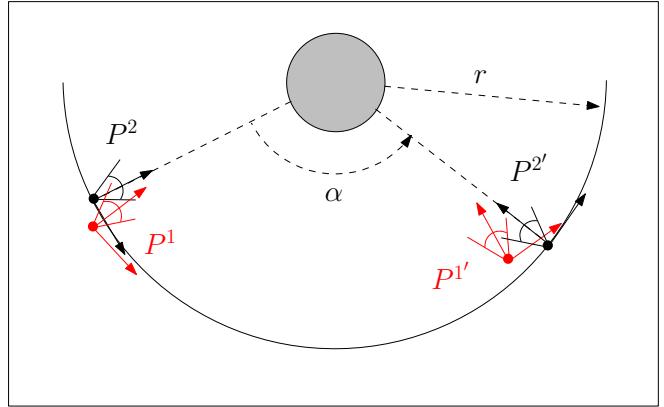
In practise, we do not compute the covariance for the 12 elements of the matrix  $P$ . Instead, we use a minimal parametrization by a 6-vector, using the Lie algebra  $se(3)$  associated to the special Euclidean group. Therefore, regarding the dimensionality of (17),  $x_i$  are 2-vectors, the Jacobians  $2 \times 6$  matrices (6 d.o.f. of the projection, and 2 coordinates of the projection), and  $\text{cov}(P)$  is a  $6 \times 6$  matrix. We restrict  $\text{cov}(P)$  to a diagonal matrix, with the diagonal elements  $\text{cov}(P)_k$ ,  $k = 1, \dots, 6$ , obtaining

$$\text{cov}(x_i) = \sum_{k=1}^6 \text{cov}(P)_k A_{i,k} \quad (18)$$

where  $A_{i,k}$  is the  $2 \times 2$  matrix

$$A_{i,k} = \begin{pmatrix} J_{Pi}(1, k) \\ J_{Pi}(2, k) \end{pmatrix} \begin{pmatrix} J_{Pi}(1, k) & J_{Pi}(2, k) \end{pmatrix}$$

<sup>3</sup> The diagonal covariance matrix is for a minimal parametrization of the projection matrix.



**Fig. 5** Camera configuration for Experiment 1. The camera matrices  $P^1$  and  $P^{1'}$  are sampled from Gaussian distributions with the means  $P^2$  and  $P^{2'}$ , respectively.  $P^2$  and  $P^{2'}$  are sampled regularly for  $r$  and  $\alpha$ .

Equation (18), together with (16), leads to

$$\sum_{k=1}^6 \text{cov}(P)_k \sum_{i=1}^n \text{tr}(A_{ik}) = \delta$$

with the six unknowns  $\text{cov}(P)_k$ , and finally we obtain:

$$\text{cov}(P)_k = \frac{\delta}{6 \sum_{i=1}^n \text{tr}(A_{ik})}, k = 1, \dots, 6. \quad (19)$$

#### 4.2 Experimental Results

*Experiment 1.* As a test model we considered 100 random points  $\{X_i | i = 1, \dots, 100\}$  uniformly sampled on the surface of a sphere. In a first experiment, we generated a set of camera pairs with different configurations, as shown in Figure 5. For each pair, the distances from the center of the sphere to the cameras were equal, and the optical axes of the cameras intersected in the center of the sphere. We restricted the analysis such that the camera centers of all configurations and the center of the sphere lie on a plane, and we denote the angle between the optical axes of the cameras  $\alpha$ . We generated 36 orientations, with  $\alpha = k\pi/36$ ,  $k = 1, \dots, 36$ . Furthermore, we considered 30 distances  $r$  between the center of the model and the center of the cameras, between 1.5 and 10.5 times the radius of the sphere. We used the same constant  $\delta$  to generate the Gaussian distribution of the perturbation. When the camera was the closest to the model, the noise was about  $10^{-4}$  of the silhouette's diagonal, and 7 times larger when the object was furthest.

For each of the  $36 \times 30 = 1080$  configurations, we generated a pair of cameras  $P^2(\alpha, r)$  and  $P^{2'}(\alpha, r)$ , computed the essential matrix  $E(\alpha, r)$  and the noise covariance matrix for each camera, constraining the deviation

of projections, as shown above, to  $\delta$ . We sampled the Gaussian distributions of the two cameras 200 times, obtaining a sequence of camera pairs  $P_i^1(\alpha, r)$  and  $P_i^{1'}(\alpha, r)$ ,  $i = 1, \dots, 200$ . We estimated the reference and epipolar distances for each  $\alpha$ ,  $r$ , and  $i$ , and the mean bias, normalized standard deviation and correlation coefficients between distances, for each  $\alpha$  and  $r$ , from the 200 test cases.

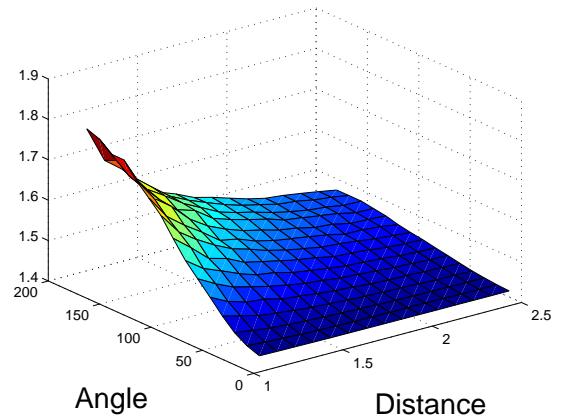
We did not observe any pattern of the ratios  $R_k$  with respect to the angle or camera-object distance. The symmetric epipolar distance had a bias of 0.95, the Sampson epipolar distance 0.66, and the manifold projection distance 0.62. The relative standard deviation  $\sigma_{rk}$  was equal to 0.3 for the symmetric epipolar distance, 0.3 for the Sampson epipolar distance, and 0.24 for the manifold projection distance. Except the cases when the epipoles were in the image and the cameras very close to the model, the symmetric epipolar distance and the Sampson distance had roughly the correlation 1.

Considering that the Sampson distance is the first order approximation of the geometric distance to the joint image manifold, which is a lower bound of the manifold projection distance (see Section 2), the experimental data suggests that the Sampson distance overestimated the geometrical distance.

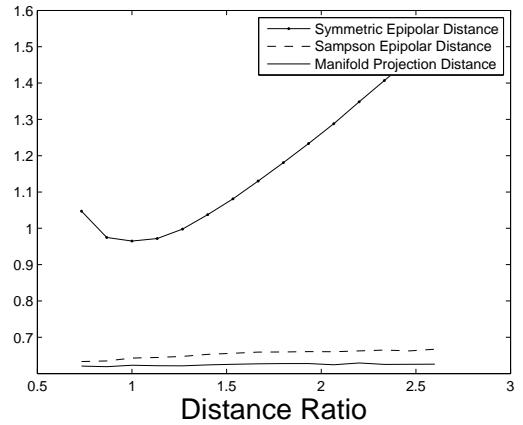
*Experiment 2.* We ran a second experiment to observe the biases of the epipolar distances under high perspective distortion, with cameras between 1.1 and 2.6 times the model's radius. The bias of the Sampson epipolar distance was dependent on the camera angle  $\alpha$  and camera-model distance  $r$ . We show all biases in Figure 6. The variation of the symmetric epipolar distance bias is hardly noticeable, as well as the variation of the manifold projection distance bias. In Figure 7 we show the average ratios between the symmetric epipolar distance and the Sampson distance.

*Experiment 3.* In a third experiment we varied the distance between the model and one camera from 1.1 to 3.9 times the model's radius, and kept the other distance at 1.5 the model's radius. We found that the bias of the symmetric epipolar distance was heavily influenced by the asymmetry of the camera distances, see Figure 8. The influence of the ratio was considerably smaller on the Sampson distance, and hardly noticeable on the manifold projection distance.

*Experiment 4.* This experiment concerned the optimization involved in manifold projection distance. In Figure 9, we show the residual and total error resulted from the manifold projection distance when the number of points used in the optimization increases from 3 to 30. The residual error is estimated only in the points used in the optimization, and the total error is estimated over all points of the model. We used a spherical model with 500 sample points. The cameras were both at a distance



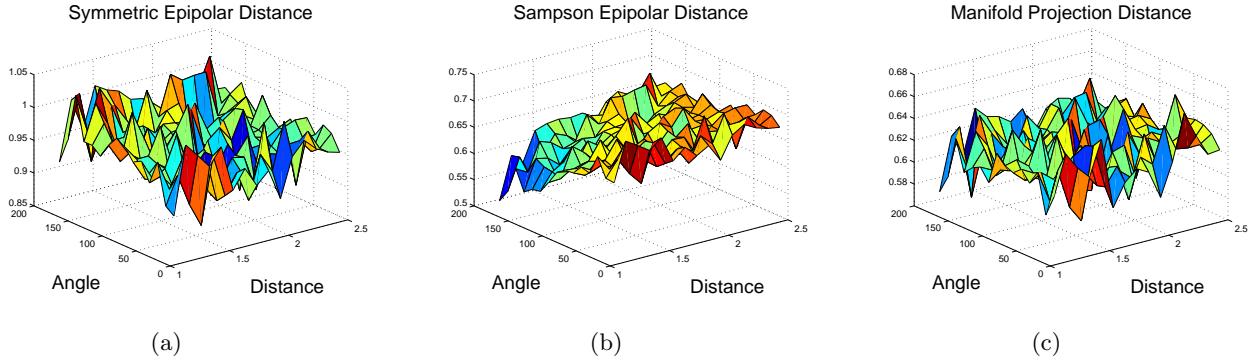
**Fig. 7** Average ratio between the symmetric epipolar distance and the Sampson distance.



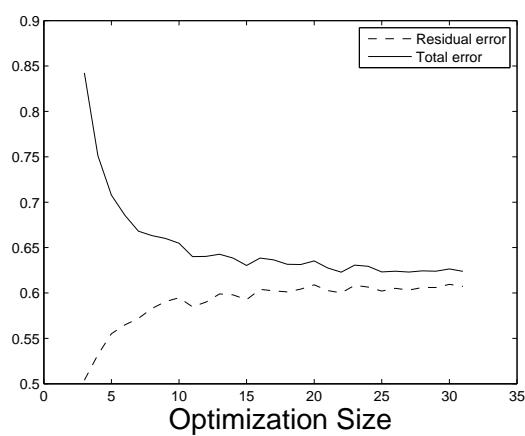
**Fig. 8** Relative bias of the epipolar distances for asymmetric model-camera distance.

of 1.5 times the model's radius, and we evaluated repeatedly several orientations. Due to the low intrinsic dimensionality of the noise model, the residual and the total error converge relatively fast.

*Interpretation.* Experimental data showed that the bias of the symmetric epipolar distance varies significantly when the distances between the cameras and the model are not equal, and, in the case of high perspective distortion, the bias of the Sampson distance varies depending of the configuration. The results seem consistent with those presented in [17, 18] with a different noise model. The manifold projection distance showed the most consistent bias with respect to the reference distance, about 0.62, close to the value found theoretically to 0.645. Re-



**Fig. 6** The relative bias of epipolar distances.



**Fig. 9** The residual and the total error of the manifold projection distance, for several optimization sizes.

garding the approximation accuracy, the unbiased manifold projection distance ranked first, as well.

## 5 Epipolar Evaluation of Texture Registration using Mutual Information

We used the evaluation scheme based on epipolar geometry to analyse the results of a texture registration method based on mutual information [5]; any other registration algorithm can be validated in the same way. Mutual information is a popular featureless registration criterion based on the assumption that aligned signals share information. The common information can be estimated numerically and used as an objective function for signal registration. Mutual information can be applied even for multimodal registration.

The registration criterion from [5] is a blend of objective functions, some defined by the mutual information between the texture intensity images and the surface

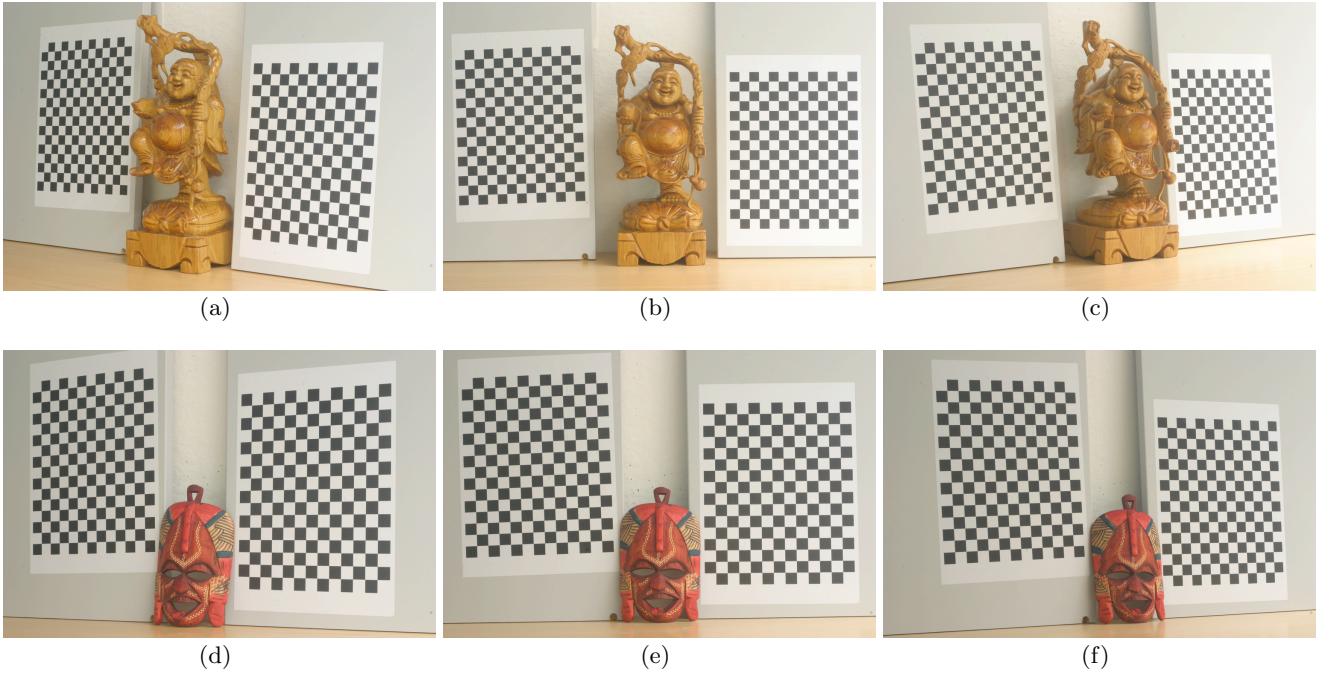
normals of a given 3D model, and others by the mutual information between the colors of image pairs. We used an improved version of [5], with joint optimization of the covariance matrices of the Parzen window, multiresolution, and the canonical weighting scheme (the objective functions are weighted equally), developed in [4]. Since the method requires initialization, we used an interactive tool based on manually selected 2D-3D correspondences.

The images were taken with a 6-megapixel digital camera (Canon EOS 10D). The camera was previously calibrated with the Bouguet toolbox, using a chessboard pattern, to find the intrinsic camera matrix, as well as the first and second order radial distortion coefficients.

To evaluate the texture registration method, we took three images with the same scene containing the object and the two patterns. We used the texture registration algorithm to obtain the extrinsic parameters of the cameras in the model's reference system, and the Bouguet toolbox to extract the extrinsic parameters in each of the patterns' reference system. We used two patterns, simultaneously visible in all images, to assess the accuracy of the gold standard. Each pattern has 180 feature points ( $12 \times 15$ ).

We performed the evaluation for two models, Shakyamuni (about 250000 points) and Mask (112000 points) [1], see Figure 10. We ran the texture registration algorithm 10 times. Each time, the parameters of the interactive registration were perturbed and used as initialization, then the optimization run. We estimated the precision using the reprojection distance, as shown in Section 1.1, and found, for each image, the values 0.33, 0.40, and 0.30 pixels for the Shakyamuni model, and 0.28, 0.27, and 0.23 pixels for the Mask model, respectively.

For each run, for each image pair and according to each pattern, we estimated the three epipolar distances. In Table 1 we show their average values over the experiment. The epipolar distances were bias-corrected, by multiplication with 1.05 (symmetric epipolar distance), 1.51 (Sampson epipolar distance), and 1.61 (manifold projection distance). The manifold projection distance was estimated using 1000 randomly chosen sample points



**Fig. 10** The images used in the experiment with Shakyamuni and Mask models. Two patterns are present in each image. The scene was not modified while taking the photographs.

Model Pattern used Image pair	Bias-corrected Epipolar Distances											
	Shakyamuni						Mask					
	left			right			left			right		
a/b	a/c	b/c	a/b	a/c	b/c	d/e	d/f	e/f	d/e	d/f	e/f	e/f
Symmetric epipolar distance	0.43	0.53	0.35	0.17	0.46	0.47	0.31	0.29	0.31	0.32	0.41	0.33
Sampson distance	0.44	0.54	0.36	0.18	0.47	0.47	0.32	0.29	0.32	0.33	0.41	0.33
Manifold projection distance	0.54	0.47	0.38	0.21	0.46	0.52	0.33	0.42	0.33	0.31	0.51	0.37

**Table 1** Epipolar distances comparing Bouguet toolbox and texture registration; two test-cases: Shakyamuni with three images, and Mask with three images. The texture registration was run 10 times and each time we computed the epipolar distances; we show the averages. Two patterns were present in each image, and the distances were computed with respect to each pattern, independently; the corresponding results are shown on the left / right sides. The distances are given in pixel units.

of the models. The minimization required by the manifold projection distance took less than one second on an AMD Athlon Dual-Core 4600+ with 2 GB of memory.

According to both assessment measures, the precision estimated with the reprojection error and the epipolar distances against camera calibration, the registration accuracy of each image was less than half-pixel. Throughout the numerical results that we found in the literature for texture registration, such as [15], this is the most accurate one, almost comparable to the one achieved by camera calibration algorithms. Though, the performance of the mutual registration registration algorithm depends on the initialization [5, 4]. The initialization accuracy that we used was typical for an interactive tool.

Each pattern covered about one third of the image area. We tried to 'balance' the scene such that the model and the patterns would fall entirely into the depth-of-field and would cover reasonable amounts of the image space. The Bouguet toolbox reported errors between one

tenth and one quarter of a pixel, causing discrepancies between the epipolar distances obtained with respect to each pattern (see Table 1). We also performed the experiment with patterns containing 64 feature points, but the discrepancies of the epipolar distances were up to 1 pixel.

The accuracy of the Bouguet toolbox is closely related to that of the corner detection algorithm, and worsens with the blur of the images. In the images we used, the scene (model and patterns) was within the depth-of-field. We attempted to take images at sharper angles with respect to the patterns, but the ground truth estimation degraded, partly because the patterns were not entirely in the depth-of-field, partly because the conditioning of the problem weakened.

A slight weakness of our approach is that both the object and the pattern (one pattern suffices) must be visible in the texture images, reducing the resolution of the object texture in the images. Of course, one may

redo the registration using images without the pattern and would expect an accuracy that is at least as good as the one measured using the images with the pattern and scaled with the corresponding resolution change.

## 6 Conclusions

We proposed a new evaluation of texture registration algorithms using, as ground truth, the epipolar geometry estimated for pairs of images. The advantage is that the ground truth can be conveniently obtained independently of geometry acquisition, requiring the object and a calibration pattern kept fixed while taking the texture images. We discussed three distance measures based on epipolar geometry: the symmetric epipolar distance, the Sampson epipolar distance, and the manifold projection distance, and analysed them in relation to the reprojection distance, standard error measure when the full ground truth is known.

We compared the distances in a synthetic experiment with ground truth, using a novel noise model. We showed that the bias of the manifold projection distance with respect to the reprojection distance can be estimated theoretically, and found the experimental values close to the expected one. We found that, after removing the bias, the manifold projection distance models the reprojection distance with the highest accuracy.

We applied the framework to evaluate a texture registration algorithm based on mutual information, against Zhang calibration, implemented in Bouguet toolbox. We found that the registration algorithm was highly accurate, below half-pixel for our case studies. The registration precision was about the same, thus, we conclude that the registration was precise and unbiased.

**Acknowledgements** This research was carried out while the first author was member of the DFG Research Training Group *Explorative Analysis and Visualization of Large Information Spaces* at the University of Konstanz, Germany, and partly while the second author was Visiting Fellow in the Department of Information Engineering, ANU College of Engineering and Computer Science at The Australian National University, Canberra, Australia. The support of these institutions is greatly acknowledged. We thank Richard Hartley, ANU, for discussions and for leading us to the theoretical analysis given in Subsection 3.1.

## References

1. 3D surface acquisition project. Website (2009). URL <http://www.inf.uni-konstanz.de/cgip/projects/surfac>
2. Bouguet, J.Y.: Camera calibration toolbox for Matlab. Website (2009). URL [http://www.vision.caltech.edu/bouguetj/calib\\_doc](http://www.vision.caltech.edu/bouguetj/calib_doc)
3. Clarkson, M.J., Rueckert, D., Hill, D.L.G., Hawkes, D.J.: Using photo-consistency to register 2D optical images of the human face to a 3D surface model. *IEEE Transactions Pattern Analysis and Machine Intelligence* **23**(11), 1266–1280 (2001)
4. Cleju, I.: Texture registration for 3D models. Ph.D. thesis, University of Konstanz (2008)
5. Cleju, I., Saupe, D.: Stochastic optimization of multiple texture registration using mutual information. In: LNCS Pattern Recognition: Proceedings Annual Symp. of the German Association for Pattern Recognition DAGM 2007, vol. 4713, pp. 517–526. Springer (2007)
6. Faugeras, O., Luong, Q.T.: *The Geometry of Multiple Images*. The MIT Press (2001)
7. Hartley, R., Sturm, P.: Triangulation. *Computer Vision and Image Understanding* **68**(2), 146–157 (1997)
8. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision*, vol. 1. Cambridge University Press (2000)
9. Leedan, Y., Meer, P.: Heteroscedastic regression in computer vision: Problems with bilinear constraint. *International Journal of Computer Vision* **37**(2), 127–150 (2000)
10. Lensch, H.P.A., Heidrich, W., Seidel, H.P.: A silhouette-based algorithm for texture registration and stitching. *Graphical Models* **63**(4), 245–262 (2001)
11. Ma, Y., Koeck, J., Sastry, S.: Optimization criteria and geometric algorithms for motion and structure estimation. *International Journal of Computer Vision* **44**, 219–249 (2001)
12. Neugebauer, P.J., Klein, K.: Texturing 3D models of real world objects from multiple unregistered photographic views. *Computer Graphics Forum* **3**(18), 245–256 (1999)
13. Seitz, S., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: Proceedings IEEE Conference Computer Vision Pattern Recognition, pp. 519–526. IEEE (2006)
14. Torr, P.H.S., Murray, D.W.: The development and comparison of robust methods for estimating the fundamental matrix. *International Journal Computer Vision* **24**(3), 271–300 (1997)
15. Troccoli, A.J., Allen, P.K.: Shadow based texture registration for 3D modeling of outdoor scenes. *Machine Vision and Applications* **18**(2), 65–72 (2007)
16. Van de Kraats, E.B., Penney, G.P., Tomazevic, D., van Walsum, T., Niessen, W.J.: Standardized evaluation methodology for 2D-3D registration. *IEEE Transactions on Medical Imaging* **24**(9), 1177–1190 (2005)
17. Zhang, Z.: Determining the epipolar geometry and its uncertainty: A review. *International Journal Computer Vision* **27**, 161–195 (1998)
18. Zhang, Z.: On the optimization criteria used in two-view motion analysis. *IEEE Transactions Pattern Analysis and Machine Intelligence* **20**(7), 717–729 (1998)
19. Zhang, Z.: Flexible camera calibration by viewing a plane from unknown orientations. In: Proceedings International Conference on Computer Vision (ICCV'99), vol. 1, pp. 666–673. IEEE Computer Society (1999)

**Ioan Cleju** is algorithms engineer at Oxford Metrics Group (Yotta DCL), United Kingdom. He received his B.Eng. from the Technical University of Iasi, Romania in 2003, his M.Sc. from the University of Joensuu, Finland in 2004, and his Ph.D. degree from the University of Konstanz, Germany in 2008. His research interests include 3D reconstruction, geometry processing, perceptual graphics, and pattern recognition.

**Dietmar Saupe** is professor of computer science at the University of Konstanz, Germany, where he chairs the research school "Explorative Analysis and Visualization of Large Information Spaces". He has a doctorate and a habilitation

degree in mathematics, both from the University of Bremen, Germany. Previous professorships were at the University of California, Santa Cruz (1985-87), and the Universities of Freiburg (1993-98) and Leipzig (1998-2002). Current areas of interest include multimedia signal processing, computer graphics, scientific visualization, and image processing.