

# Highly-Automatic MI Based Multiple 2D/3D Image Registration Using Self-initialized Geodesic Feature Correspondences

Hongwei Zheng, Ioan Cleju, and Dietmar Saupe

Computer and Information Science, University of Konstanz  
Fach M697, 78457 Konstanz, Germany  
{firstname.surname}@uni-konstanz.de

**Abstract.** Intensity based registration methods, such as the mutual information (MI), do not commonly consider the spatial geometric information and the initial correspondences are uncertainty. In this paper, we present a novel approach for achieving highly-automatic 2D/3D image registration integrating the advantages from both entropy MI and spatial geometric features correspondence methods. Inspired by the scale space theory, we project the surfaces on a 3D model to 2D normal image spaces provided that it can extract both local geodesic feature descriptors and global spatial information for estimating initial correspondences for image-to-image and image-to-model registration. The multiple 2D/3D image registration can then be further refined using MI. The maximization of MI is effectively achieved using global stochastic optimization. To verify the feasibility, we have registered various artistic 3D models with different structures and textures. The high-quality results show that the proposed approach is highly-automatic and reliable.

## 1 Introduction

Multiple 2D/3D image registration and mapping is a key problem in computer vision that shows up in a wide variety of applications such as medical image analysis, object tracking, recognition and visualization. In practice, due to the less information about intra- and inter-correspondences for captured multiple images and the 3D model, the problem of multiple 2D/3D image registration is highly ill-posed. The MI measure based 2D/3D image registration methods take only intensity values into account without considering spatial geometric information. The error of initial correspondences may easily lead to a blunder in the final registration. Therefore, it is often that the initial spatial correspondences are manually determined by users, which is non-efficient and time consuming.

This paper presents a novel approach for highly-automatic MI based texture registration using self-initialized geodesic feature correspondences. Given a 3D shape model and multiple images, we perform this approach in three main steps with respect to the search of initial spatial correspondences, the estimation of projective transformation of multiple views, and the refinement of texture registration. At first, we extract local features of surfaces on a 3D shape. However,

direct geometric feature extraction on 3D shapes is difficult. The reason is that the scale-variability of geometric structures on scanned 3D models are simplified due to discrete 3D coordinate point-clouds or triangulated surfaces on 3D models, shown in Fig. 1. Traditional work extract feature descriptors using only surface and curvature smoothing based on 3D coordinates lacking canonical scale analysis. The important discriminative information encoded in the scale-variability of intrinsic geometric structures are easily ignored. Recent work on the scale-variability of images as a 2D projection of 3D objects [1], scale invariant features in the discrete scale space for 2D images [2], and scale-dependent 3D geometric features [3] have been studied intensively. Inspired from these work, we present a comprehensive framework for analyzing and extracting local feature descriptors using the constructed multiple normal maps in geometric scale-spaces.

Secondly, the projective transformation of correspondences for image-to-image 2D/2D and image-to-model 2D/3D are estimated using sparse geometric features and related camera parameter estimation procedures. The key idea underlying the correspondence between a 3D shape and images is that multiple 2D normal images are the 2D projections of geometric surfaces on a 3D shape. The converted multiple normal maps encode the rich geometric information within the spatial distribution of each local features that are sparsely distributed on the normal images using geodesic distance measure. The self-initial correspondences are estimated in two steps with respect to 2D/2D and 2D/3D correspondences. Furthermore, a maximization of MI method [4], [5] is extended for refining multiple 2D/3D image registration using self-initialized geodesic feature correspondences.

Our approach has several advantages. First, the approach is highly-automatic and efficient, facilitating the human supervision and the search of initial correspondences to geometric 3D models with varying geometric complexity. One only needs to define the group of input images according to their surface representations on the 3D shapes. The best position of given multiple viewed images to the 3D surface is determined using self-initialized correspondences. Second, sparse geometric features based initial correspondence does not require camera captured images to contain the entire 3D object for the purpose of silhouette extraction [6], or shape outline extraction. Third, the approach allows sparse geometric feature correspondences and entropy MI based optimization to be integrated for solving a practical problem in a reliable and optimal way.

The rest of the paper is organized as follows. Section 2 presents the concepts of multiple normal images of a 3D shape model. Section 3 describes the geodesic measure and sparse geometric feature extraction. Section 4 formulates the global stochastic optimization based maximization of MI. Implementation details and experimental results are presented in Section 5.

## 2 Multiple Normal Image Representation of a 3D Model

Given a triangular mesh model of a 3D object, a original mesh  $M$  is parameterized to a planar domain  $D$ . The parametrization  $p : D \rightarrow M$  is a bijective mapping from a discrete set of planar points to the mesh vertex set. Normally,

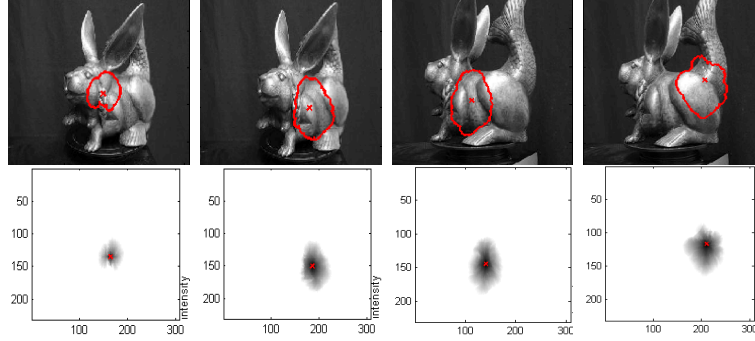


**Fig. 1.** *a|b|c|d.* (a) A partial point-clouds 3D model. (b) Shaded 3D model. (c) Normals on the 3D model. (d) 2D mesh parameterization with fixed boundary of this 3D model.

parameterizing a mesh to a planar domain does not preserve the angles and the surface area of faces on the mesh. Some changes in the angles and surface area are considered as distortions. Minimizing the parameterization distortion is a challenging problem [7], [8], [9]. For example, in Fig. 1(d), we parameterize one 2D mesh image that encodes the original 3D model.

Although the geometric properties can be encoded via 3D coordinates and curvatures, the surface normals have been demonstrated as a suitable base representation, shown in Fig. 1(c). It allows us to use Gaussian filtering on the 3D shape without influencing the topology of 3D shapes. The normal directions are critical for detecting 3D features. To implement it, we project the surface normal of each vertex of the 3D model and then interpolate those values in the planar domain using barycenter coordinates within each triangular face to obtain a dense normal map. The resulting normal map is a geometric 2D image representation of the original 3D shape (or part of the shape) that is independent of the resolution of its 3D model. To achieve accurate representation of geometric surfaces on a 3D shape to one or multiple normal images, we define the transformation in terms of geodesic distances instead of the traditional Euclidean distances.

The generated one or multiple normal maps cannot exactly represent the original 3D shape due to the distortion of parameterization. For example, the distance between any two points in the normal map is not equivalent to the corresponding relative geodesic distance on the 3D model. To construct 2D surface representation of the original shape, the correct relative geodesic distances between any two points on the normal map is necessary. Therefore, the distortion is computed for each point in the normal map. Given a point  $v \in D$  that maps to a 3D mesh vertex  $\Psi(v)$ , we define its distortion  $\xi(v)$  in the equation  $\xi(v) = \frac{1}{|adj(v)|} \sum_{u \in adj(v)} \frac{\|v-u\|}{\|\Psi(v)-\Psi(u)\|}$ , where  $adj(v)$  is a set of vertices adjacent to  $v$ . The local distortion is a measure of the average change in the length of the edge adjacent to a vertex. The large  $\xi(v)$ , the more the adjacent edges have been stretched in the parameterization around  $v$ . We then construct a dense map of distortion values in this way. The resulting distortion map is to approximate the geodesic distances between any two points in the normal map.



**Fig. 2.**  $\frac{a|b|c|d}{e|f|g|h}$ . Geodesics Gaussian distribution on embedded surfaces in 3D spaces. (a)(b)(c)(d) The boundary of geodesic Gaussian kernel with a radius. (e)(f)(g)(h) The distribution of geodesic gaussian kernels with a given radius.

### 3 Geodesics Measure and Feature Corners Extraction

To achieve initial correspondence searching, we have to define the right feature spaces. To detect salient feature corners for matching, we first derive the first order and second order partial derivatives of the normal map  $N_\sigma$ . Novel feature corners can then be derived using these partial derivatives using geodesics distance as the distance metric which can accurately represent the local surface geometry in scale-space. Given a 2D isotropic Gaussian centered at a point  $u \in D$ , we define the value of geodesic Gaussian kernel at a point  $v$ , then the boundary of geodesic support region on a intensity surface in 3D space. The geodesic Gaussian kernel is derived as  $G_{geod}(v, u, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{d_{geod}(v, u)^2}{2\sigma^2}\right]$ , where  $d_{geod} : R^2 \times R^2 \rightarrow R$  is the geodesic distance between the 3D surface points  $\Psi(v)$  and  $\Psi(u)$ . The geodesic distance between two 3D points is defined as the discretized line integral  $d_{geod}(v, u)$  in the distortion map, which can be computed as  $d_{geod}(v, u) \approx \sum_{v_i \in \Omega(v, u)} \frac{\xi(v_i)^{-1} + \xi(v_{i+1})^{-1}}{2} \|v_i - v_{i+1}\|$ , where  $\Omega(u, v) = [v_1, v_2, \dots, v_n, u]$  is a list of points sampled on the surface between  $v$  and  $u$ . The density of this geodesic sampling determine the quality of the approximation of the original geodesic distance. Using the geodesic Gaussian kernel, the normal at point  $u$  for scale level  $\sigma$  as  $N_\sigma(u) = \sum_{v \in F} N(v)G_{geod}(v; u, \sigma) / \|\sum_{v \in F} N(v)G_{geod}(v; u, \sigma)\|$ , where  $F$  is a set of points in a window centered at  $u$ . The window size is also defined in terms of geodesic distance and is proportional to the standard deviation  $\sigma$  at each scale level. In our implementation, we change the size of the window from the center point while evaluating each point's geodesic distance from the center to correctly estimating the distribution of similar "high" points. Note that the geodesic Gaussian kernel can be performed for the image embedded surface with the 3rd coordinate of intensity in 3D space. Fig. 2 shows the non-isotropic boundary distribution of geodesic Gaussian kernel with a scale  $\sigma$  for an embedded surface in 3D spaces. It has the same effects on the normal maps which can be smoothed using the geodesic Gaussian kernel in different scales.

Based on these properties, we extract the geometric meaningful feature corners on the normal maps so that we can find the correspondence points between normal maps and the photographs of the real object in the next step. We are interested in detecting the geometric corners which can also be detected on the related 2D photographs. Firstly, we compute the Gram matrix  $\mathcal{R}$  of the first order partial derivatives of the normal map  $N_\sigma$  at each point. The Gram matrix at a point  $u$  is defined as

$$\mathcal{R}(u; \sigma) = \sum_{v \in W} \begin{bmatrix} N_\sigma^s(v)^2 & N_\sigma^s(v)N_\sigma^t(v) \\ N_\sigma^s(v)N_\sigma^t(v) & N_\sigma^t(v)^2 \end{bmatrix} G_{geod}(u, v) \quad (1)$$

Where  $N_\sigma^s$  and  $N_\sigma^t$  are the horizontal first derivatives and vertical first derivatives in the tangential plane on the normal map. The normal map encodes rich geometric information parameterized from the 3D model. Then, using a principle of corner extraction [10], we extract corners from normal maps using the likelihood of a corner as the corner response following  $\mathcal{P}(u; \sigma) = \det(\mathcal{R}(u; \sigma)) - \tau \text{Trace}(\mathcal{R}(u; \sigma))^2$ , where  $\tau$  is a tunable parameter. Once the corners are detected in 2D they can be mapped back to the original 3D model. Since the 2D normal map is dense, the corresponding location of the corners in 3D are independent of the input 3D model's triangulation.

#### 4 Maximization of MI Based Registration Refinement

Let  $x$  be an arbitrary point on the surface of the 3D model that is visible in the texture image  $I(x)$ , and  $T(x)$  is the 3D-2D projective transformation. The intensity value in the image  $I(T(x))$  depends on the radiance in the scene, the (bidirectional reflectance distribution function) BRDF of the surface in  $x$ , and the normal to the surface  $N(x)$ . The dependance of the intensity  $I(T(x))$  on the surface normal  $N(x)$  can be estimated using the MI (redefining  $x$  as a random variable), i.e., the entropy of the “normal-intensity”  $\mathcal{M}[N(x), I(T(x))]$ . If we consider a random variable  $x$ , its entropy  $\mathcal{H}(x)$  can be estimated from two independent samplings of  $x$ . One sampling is used to estimate the probability density function with the Parzen window method, which is then evaluated on the second sampling. The MI between  $N(x)$  and  $I(T(x))$  is estimated from small sub-samplings of the data (order of tens of points). When the image  $I(T(x))$  and the model  $N(x)$  are optimally aligned, the mutual information  $\mathcal{M}$  is maximized.

Based on this principle, Viola and Wells [4] proposed a method based on Parzen windows [11] to estimate the MI and its differential with respect to  $T(x)$ ,

$$\mathcal{M}[N(x), I(T(x))] = \mathcal{H}[N(x)] + \mathcal{H}[I(T(x))] - \mathcal{H}[N(x), I(T(x))] \quad (2)$$

where  $\mathcal{M}$  is MI and  $\mathcal{H}$  entropy. Note that  $\mathcal{H}[N(x)]$  does not depend on the transformation  $T(x)$ . Due to the small random sub-sampling of data, the estimation of the MI gradient is stochastic. Recently, Cleju and Saupe [5] extended this formulation to support multiple image registration. Except the image-to-model MI which is defined in the same way as [4], additional image-to-image MI objective

function are defined using the full color space. The MI between images  $I_1$  and  $I_2$ , associated with the transforms  $T_1$  and  $T_2$ , is defined as:

$$\mathcal{M}[I_1(T_1(x)), I_2(T_2(x))] = \mathcal{H}[I_1(T_1(x))] + \mathcal{H}[I_2(T_2(x))] - \mathcal{H}[I_1(T_1(x)), I_2(T_2(x))]$$

The main advantage of the formulation is that the new objective functions bring the surface BRDF in the global optimization criterion. In this function, the image-to-image MI is defined from the chrominance components  $I$  and  $Q$  of the YIQ color space. The image-to-image MI is parameterized by the projective transformations associated with both images, and it is maximized when both images are aligned to the model. The stochastic gradient estimation follows the same procedure as for the image-to-model MI.

For the joint refinement and optimal registration of several images to a 3D model, we use a multi-objective optimization method [5] that is defined as a linear combination of the elementary MI functions with non-negative weights. In this method, firstly, when all images are aligned and the registration refined to the model, all objective functions are maximized. If only the image-to-model MI functions are considered, each set of camera parameters corresponds to one objective function. In this case the iterative gradient-based optimization updates each set of parameters in the direction of the corresponding gradient. Secondly, when all image-to-image MI functions are considered, we estimate several gradients for the parameters of each camera, corresponding to the MI with the model and with other overlapping images. In each iteration we must choose the direction for optimization based on the these gradients.

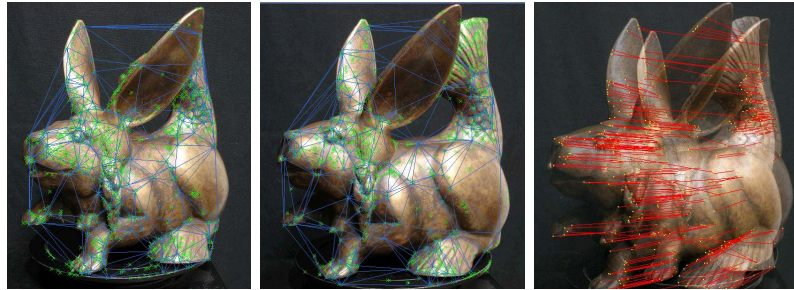
## 5 Implementation and Experimental Results

We present the implementation in detail. Sequentially, we describe the intermediate and final registration results in our experiments for the proposed approach.

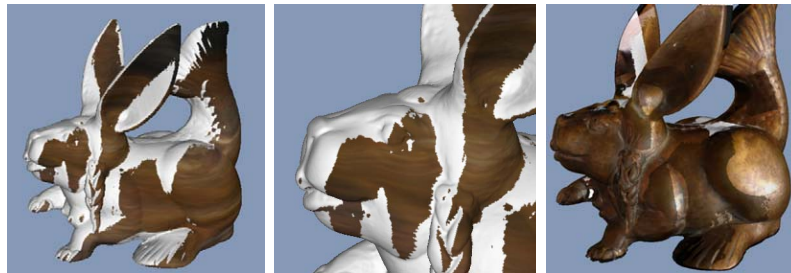
### 5.1 Self-initialized Correspondence Using Sparse Features

In this step, we estimate initial correspondence for both image-to-image and image-to-model matching. We have implemented the proposed geodesic feature corner extraction method on the normal maps which is parameterized from the 3D shape models. Then we present automatic initial correspondence results for image-to-image and image-to-model via certain related matching algorithms and a camera self-calibration method. Finally, we further optimize the extrinsic parameters of the cameras and refine the registration results using the MI based global stochastic optimization framework.

Firstly, the projective transformation  $T$  is computed by estimating camera model. The optimization model does not make assumptions on the projective transformation  $T$ , and consequently on the camera model. In our implementation we consider the pinhole camera model [12] with four distortion coefficients (two for radial distortion and two for tangential distortion). Since multiple views of photographs have been taken around the real 3D object, any two uncalibrated



**Fig. 3.** *a|b|c* Initial correspondence searching of pairwise texture images. (a)(b) Two captured texture images for the real 3D object with detected and triangulated feature corners. (c) Unsupervised searching of initial feature correspondence between pairwise texture images (a) and (b) using the GASAC method.



**Fig. 4.** *a|b|c* (a) Project two images to the 3D model using the self-initialized feature correspondence. (b) Zoom in (a). (c) After projective transformation, these texture images are initially mapped to the surface of the 3D model. Note that these two images were taken with different illuminations (one texture image was taken using flashlight).

image consequences can be used to estimate the fundamental camera matrix and epipolar lines for pairwise image-to-image matching [13]. The intrinsic parameters field-of-view, optical center, and distortions, were self-calibrated using Zhang's method [14] which are used as initial value for constructing image-to-model correspondence. Any of projective transformation parameters are further refined and optimized using the MI based multiple objective functions. In our experiments, we consider the intrinsic parameters fixed and we optimize only the extrinsic parameters. The rotation matrix is parameterized by axis-angle form for its advantages over Euler angles in the iterative optimization [15].

Secondly, image-to-image 2D/2D correspondences are self-initialized using an extended RANSAC algorithm. The initial correspondence for 2D/2D matching is to determine the relative orientation of the images. The self-initialization is estimated using the methods in multiple view geometry [13], [12]. We have estimated the fundamental matrix and the epipolar lines for describing the projective relative orientation of uncalibrated images. To handle the large number of highly resolving images, the computationally intensive RANSAC algorithm for robust



**Fig. 5.**  $\frac{a|b|c|d}{e|f|g|h}$  High-fidelity 2D/3D texture registration results using the suggested method. (a)(e) Camera photo for the real 3D object. (b)(f) The scanned 3D model. (c)(g) and (d)(h) Final results of multiple image registration on the 3D model.

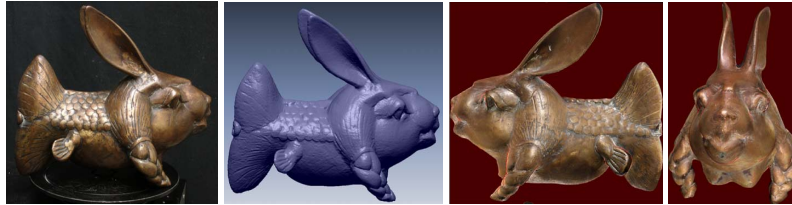
outlier detection has been replaced by a faster evolutionary approach called Genetic Algorithm Sampling Consensus GASAC [16]. The initial correspondence value and results of the feature based matching are shown in Fig. 3.

Furthermore, image-to-model 3D/2D correspondence are estimated based on estimated camera parameters and additional camera calibration. To find initial correspondence value between the 3D shape and the given texture images, we have extracted and utilized the obtained geodesic feature descriptors using normal maps of the 3D shape. We convert the 3D/2D matching problem to a 3D/3D one. Similar to the ICP algorithm [17], the search of initial correspondence is to estimate a paired-point matching transformation based on geodesic feature points and then apply the transformation to one side of the registration. We have scanned several artistic 3D objects with metal surface and reflection. In Fig. 4(a) and (b), we can see some images are initially projected to the surface on the 3D model without any projective transformation processing using sparsely distributed correspondent feature points. After the projective transformation using self-calibrated transformation parameters, shown in Fig. 4(c), several images with different illuminations are mapped on the surface of the 3D shape model after refined projection and registration.

## 5.2 Color Image Registration and Comparison of Other Approaches

The multiple image registration can be refined and tuned using the maximization of mutual information. If a patch of the surface is visible in two images after self-initialized correspondence, we will simply say that the images overlap. We define image-to-image MI functions for each overlap. In [18], a 3D model with reflectance values mapped on its surface was registered to color images using the method [4]. Here, we extend the objective function for fully color information





**Fig. 6.** *a|b|c|d* High-fidelity 2D/3D texture registration results using the suggested method. (a) Photograph. (b) The scanned 3D model. (c)(d) Final results of multiple image registration using the suggested method. Look from different viewpoints.

of the images to the registration objective functions. In our implementation we defined the image-to-image MI from the chrominance components I and Q of the YIQ color space. The image-to-image MI is parameterized by the projective transformations associated with both images, and it is maximized when both images are aligned to the model. The gradient estimation follows the same procedure as for the image-to-model MI. Fig. 5 and Fig. 6 show the refined final registration results using the suggested approach.

In the refinement period of registration, the maximization of MI does not need the existence of any 3D-2D feature information (including visible outlines in the image) after self-initialized correspondence. The refinement of global stochastic optimization does not make assumptions on the unknown parameters of the rendering function. It is robust to various illumination conditions and even to occlusions. Compare to other registration algorithms, our accuracy is significantly better than the one reported in [19], mean projection error of 5-6 pixels for  $3072 \times 2304$  pixel images. One restriction when using the MI objective function is that the value of the global maximum cannot be estimated. In contrast, when registration is refined and further optimized with point correspondences, for instance, the global optimum corresponds to 0 projection error. The whole software system has been implemented in Cpp language based on SUSE 10.3 Linux system. In Fig. 5, the full registration and refinement of six images takes 10 minutes using an AMD Athlon Dual-Core 4600+ 2GB PC.

## 6 Conclusions

In this paper, we have suggested a new approach for reliable and highly-automatic multiple image registration using self-initialized geodesic feature correspondences. To improve the robustness and automation of MI based multiple 2D/3D image registration, the integration of self-initialized geodesic feature correspondence can support accurate initial value for further global stochastic optimization based maximization of MI. Furthermore, we have demonstrated that the normal maps of the 3D model in scale space encode rich geometric information for geodesic feature extraction. In particular, we have combined and utilized both advantages from sparse features and entropy MI in an integrated framework. A thorough evaluation and several high-fidelity registered 3D point based models show that

the proposed approach has more flexibilities for highly-automatic and reliable multiple image registration.

## Acknowledgments

The work was supported by the DFG GK 1042 “Explorative Analysis and Visualization of Large Information Spaces” at the University of Konstanz in Germany.

## References

1. Witkin, A.P.: Scale-space filtering. In: Proc. Int. Joint Conf. Artif. Intell., IJCAI Karlsruhe, pp. 1019–1021 (1983)
2. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), 91–110 (2004)
3. Novatnack, J., Nishino, K.: Scale-dependent 3D geometric features. In: IEEE Eleventh International Conference on Computer Vision (2007)
4. Viola, P.A.: Alignment by maximization of mutual information. Technical Report AITR-1548 (1995)
5. Cleju, I., Saupe, D.: Stochastic optimization of multiple texture registration using mutual information. In: Hamprecht, F.A., Schnörr, C., Jähne, B. (eds.) DAGM 2007. LNCS, vol. 4713, pp. 517–526. Springer, Heidelberg (2007)
6. Lensch, P., Heidrich, W., Seidel, H.P.: Automated texture registration and stitching for real world models. In: Pacific Graphics (2000)
7. Floater, M.S., Hormann, K.: Surface parameterization: a tutorial and survey, pp. 157–186. Springer, Heidelberg (2005)
8. Yoshizawa, S., Belyaev, A., Seidel, H.-P.: A fast and simple stretch-minimizing mesh parametrization. In: Int. Conf. on Shape Modeling and Applications (2004)
9. Eck, M., DeRose, T., Duchamp, T., Hoppe, H., Lounsbery, M., Stuetzle, W.: Multiresolution analysis of arbitrary meshes. In: ACM SIGGRAPH 1995, pp. 173–182 (1995)
10. Förstner, W.: A feature based correspondence algorithm for image matching. *Int. Arch. Photogrammetry Remote Sensing* 26(3), 150–166 (1986)
11. Duda, R.O., Hart, P.E.: *Pattern Classification*, 2nd edn. Wiley Interscience, Hoboken (2000)
12. Faugeras, O.: *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, Cambridge (1993)
13. Hartley, R.I., Zisserman, A.: *Multiple View Geometry in Computer Vision*, 2nd edn. Cambridge University Press, Cambridge (2004)
14. Zhang, Z.: Parameter estimation techniques: A tutorial with application to conic fitting. *Journal of Image and Vision Computing* 15(1), 59–76 (1997)
15. Taylor, C.J., Kriegman, D.J.: Minimization on the lie group  $SO(3)$  and related manifolds. Technical report, Dept. of E.E. Yale University (1994)
16. Rodehorst, V., Hellwich, O.: Genetic algorithm sample consensus (GASAC) - a parallel strategy for estimation. In: 25 Years of RANSAC in CVPR (2006)
17. Penney, G.P., Edwards, P.J., King, A.P., Blackall, J.M., Batchelor, P.G., Hawkes, D.J.: A stochastic iterative closest point algorithm. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, p. 762. Springer, Heidelberg (2001)
18. Nishino, K., Sato, Y., Ikeuchi, K.: Appearance compression and synthesis based on 3D model for mixed reality. In: IEEE Int. Conf. Computer Vision, pp. 38–45 (1999)
19. Jank, Z., Chetverikov, D.: Photo-consistency based registration of an uncalibrated image-pair to a 3D model using genetic algorithm. In: 3DPVT, pp. 616–622 (2004)