A Robust Multiresolution Registration Approach

Stephan Seeger and Gerd Häusler

University of Erlangen, Chair for Optics
Staudtstrasse 7/B2, D-91058 Erlangen, Germany
Email: {seeger,haeusler}@physik.uni-erlangen.de

Abstract

The task of registering several range images taken from different viewpoints into a single coordinate system is usually divided into two steps: first, finding a rough estimate of the searched transformation on the base of reduced parts of the data sets, and second, finding the precise transformation with another method on the base of the complete information given by the data sets. Using different approaches for these registration steps has no fundamental but only practical reasons: for the fine tuning step it exists an easily implementable algorithm whose efficiency and precision has been proven in many experiments, but which is not applicable for finding an initial rough estimate of the searched transformation. We present a multiresolution approach to the registration problem that has the potential to combine these two registration steps and is based on hierarchical Hough methods and local frames defined in each surface point.

1 Introduction

In order to digitize the surface of a complex object, several range images have to be taken from different viewpoints. To generate a complete model from these images, in a first step all of them have to be transformed to the same coordinate system. In order to find the rigid transformation (rotation + translation) between two range images with overlapping surface regions usually the standard *Iterated Closest Point* algorithm (ICP) [1, 4, 23, 8, 20, 19] is performed. The ICP algorithm iteratively identifies closest points as corresponding points and then minimizes in each iteration step a least squares sum of the distances between these corresponding points with respect to the transformation parameters. Since the whole information of the images can be used, such an approach allows a quite precise registration. However, the algorithm has the drawback that it only converges to the correct transformation if it starts with a good estimation of the searched transformation. To get this initial transformation, often features are extracted from the data and then attached to each other, either manually [15] or by using complicated heuristics [20, 8, 5, 21, 22]. Other approaches try to find a good estimation of the correct transformation by global optimization strategies like genetic algorithms [2] or mean field theory [17]. To summarize, all known registration approaches divide the registration task into two steps: first finding a rough estimate on the base of a reduced data set, then changing the method and finding the precise transformation based on the whole given information. We believe that some aspects of this general approach are inevitable: firstly, the initial estimation of a rough transformation is only possible on a reduced data set because of the complexity of the problem, and secondly, the most precise transformation can only be determined by using the whole information of the data sets. However, in principle there is no reason to change the registration method between these two steps. Therefore we have developed a multiresolution approach to the registration problem that has the potential to combine the two registration steps. The new approach is based on hierarchical Hough methods and local frames defined in each surface point. Hough methods [16, 9] as well as local frames [8] have been already used for finding rough estimates of transformations. Hierarchical Hough methods are also well known [11]. We show that their combination in a multiresolution ansatz promises a unified approach to the rigid registration problem.

2 Multiresolution Registration

2.1 Basic Idea

In a first step of our algorithm we calculate a local frame for every data point of both range images. With the help of these local frames it is possible to calculate for each point in range image 1 the transformation to each point in range image 2. The transformation parameters $(\alpha,\beta,\gamma,t_x,t_y,t_z)$ are stored in a six dimensional accumulation table by incrementing a counter in the table at position $(\alpha,\beta,\gamma,t_x,t_y,t_z)$. Methods using accumulation tables in this fashion are called Hough methods. Since transformations calculated from correct point correspondences result in the same transformation while all other transformations are distributed more or less randomly in the parameter space, we can expect a peak in the Hough table at the position of the searched transformation (see Figure 1).

2.2 Problems

Transforming each point of image 1 in each point of image 2 is an $O(n^2)$ operation (n is the number of points in each image), which cannot be performed in reasonable time when n is large as is the case for a typical range image ($n \approx$ 100.000). In addition the calculation of local frames (and therefore also the calculation of the transformations) may be heavily influenced by noise, especially if we use differential quantities for this computation. Last but not least the memory consumption of the six dimensional Hough table can be prohibitive. For example, if we wanted an angle accuracy of 5° we would need $360°/5° = 72$ intervals in this parameter dimension of the Hough table. Using just as many intervals in each parameter dimension we would

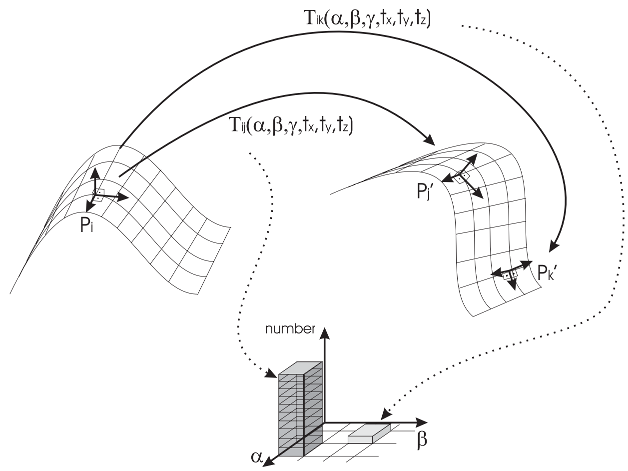

Figure 1: Each pair of points (P,P') with one point from the first image and one point from the second image defines a transformation with the help of their local frames. The correct transformation between the images is given by the most frequently occuring transformation in the Hough table.

get $72^6 \approx 140 \cdot 10^9$ cells in the Hough table. At each cell we have a counter that needs at least 2 bytes so that the total memory consumption of the Hough table under the above conditions would be at least 280 GB.

2.3 Solutions

In order to solve the $O(n^2)$ problem, we have to reduce n. This can be done by transforming regions instead of points into each other. The local frames we need for the determination of the transformation can be averaged over all points in the regions. There is also the hope of reducing the influence of noise of the local frames by this averaging process.

In order to reduce the memory consumption of the Hough table there are two possibilities. Firstly, instead of using one six dimensional Hough table we can use two 3-dim, three 2-dim or six 1-dim Hough tables that need significantly less memory to reach the same accuracy for the determination of the transformation. Secondly, we can limit the number of intervals in each Hough table dimension so that the memory consumption of the whole Hough table is acceptable. To reach a high accuracy in the Hough ta-

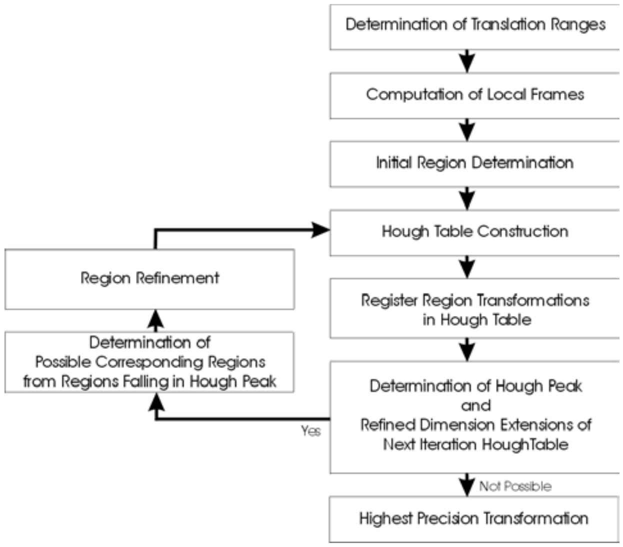

Figure 2: The multiresolution registration algorithm.

ble we can refine the subdivision of the Hough table after detection of a Hough peak and then iterate this procedure. Since one 6-dim Hough table promises the best signal to noise ratio we decided to use the hierarchical approach for a 6-dim Hough table. In addition, in this way the accuracy in the determination of the transformation is not limited by memory considerations.

2.4 The Algorithm

For an overview of the algorithm see Figure 2.

2.4.1 Initialization

Determination of Translation Ranges. In a first initialization step we limit the range of reasonable translations: after an initial translation overlaying the centers of mass of the two range images, the norms of all reasonable remaining translations are bounded by the sum of greatest distances of points in image 1 and of points in image 2 to the joint center of mass.

Computation of Local Frames. In a second initialization step we have to compute local frames for every point of both range images. Using a differential geometry approach we can compute principal frames consisting of the normal in the given point and the directions of minimal and maximal curvatures [6, 8]. Usually these directions are calculated from the eigenvectors of the Weingarten map with the help of a local polynomial approximation of the surface [13]. However, more efficient computation schemes have already been developed [12]. Another possibility to compute local frames that does not rely on differential quantities is described in [5].

Initial Region Determination. In a third initialization step we have to determine initial regions that are transformed into each other with the help of averaged local frames. This should be done without introducing threshold values since optimal values usually depend on the given object and the threshold value optimization is a separate problem. In addition it should be possible to predict the number of resulting regions of this segmentation step since otherwise the whole algorithm can fail due to the $O(n^2)$ complexity. Furthermore, the resulting regions in two range images from different viewpoints should approximately correspond to each other so that the transformations of all corresponding regions are approximately the same and therefore can accumulate in the Hough table. Last but not least it should be possible to refine the resulting regions to allow taking more and more information into account during the iterations of the algorithm.

We propose a curvature based 2D Haar wavelet reduction for the region determination and a quadtree decomposition for an elegant and efficient access to the averaged local frames of these regions. Details can be found in Subsection 2.5. Our approach relies on the raster structure of the range images. It should be possible to generalize our approach to triangle meshes by using a progressive mesh (PM) representation [10, 3] or by a multiresolution analysis of the surfaces with the help of a surface wavelet construction [18, 7]. However, one advantage of our approach is the possibility of a straightforward generalization to volume data.

2.4.2 Iteration

After the initialization steps we have to iterate the following steps:

Hough Table Construction. Firstly, a 6-dim Hough table is constructed. The extensions of each dimension are determined by the last iteration step. The initialization of the translation dimensions are determined by the first initialization step. When using Euler angles for the representation of the rotation the initialization of the rotation dimensions are given by $\alpha = 0 \ldots 360^\circ, \beta = 0 \ldots 180^\circ, \gamma = 0 \ldots 360^\circ$. The number of Hough table cells and therefore the resolution of the Hough table follows from signal to noise ratio considerations described in Subsection 2.6.

Region Transformations. Secondly, regions from both range images are transformed into each other resulting in the following rotation $\mathbf{R}$ and translation $\mathbf{t}$:

$$\mathbf{R}_{i \to j'} = \mathbf{R}_{j'} \mathbf{R}_i^t, \tag{1}$$
$$\mathbf{t}_{i \to j'} = \mathbf{p}_{j'} - \mathbf{R}_{j'} \mathbf{R}_i^t \mathbf{p}_i \tag{2}$$

where $\mathbf{p}_i$ is the center of mass of region i and the columns of $\mathbf{R}_i$ are given by the averaged local frames of region i. After each calculation of a transformation a counter in the Hough table at the position of the transformation parameters $(\alpha, \beta, \gamma, t_x, t_y, t_z)$ is incremented. It can be incremented just by one or in a more sophisticated way in dependence of differences of expected invariances of the regions. As invariances can be taken e.g. averaged minimal or maximal curvatures, the areas of the regions or, when using additional information like the texture, the differences of the intensity values to the neighboring regions. Therefore, the counter $c_{\alpha,\beta,\gamma,t_x,t_y,t_z}$ at the Hough table position $(\alpha, \beta, \gamma, t_x, t_y, t_z)$ can be updated by

$$c_{\alpha,\beta,\gamma,t_x,t_y,t_z} + \frac{1}{1 + \sum_k \left(a_i^{(k)} - a_{j'}^{(k)}\right)^2} \to c_{\alpha,\beta,\gamma,t_x,t_y,t_z} \tag{3}$$

where $a_i^{(k)}$ is the k-th invariant attribute of region i. The term $\frac{1}{1+(\cdot)}$ ensures that a transformation has the more influence, the less is the difference between the invariant attributes of the regions, and that the increment is limited to one.

Hough Peak Determination. Thirdly, the maximal entry in the Hough table, i.e. the Hough peak, has to be detected. Since transformations that lie on the border of a Hough table cell can be accidently distributed over many cells, in a first step each counter in the Hough table is summed over all counters of its direct neighboring cells before determining the Hough peak. When the Hough peak is detected we also have the refined dimension extensions for the next iteration Hough table: we take the direct neighbors of the Hough peak cell as the new ranges. If, due to a too small signal to noise ratio, no Hough peak can be detected in the highest region resolution, the best possible transformation is found and the algorithm stops.

Determination of Possible Corresponding Regions. Fourthly, in order to avoid transforming again each region of image 1 into each region of image 2 in the next iteration step and in order to allow in this way a region refinement without failing on the $O(n^2)$ problem, possible corresponding regions have to be determined from transformations falling in the Hough peak cell. To analyze this, each region of image 1 has again to be transformed into each region of image 2. Whenever a transformation falls in the new range of the next iteration Hough table, the appendant regions are memorized as possible corresponding regions.

Region Refinement. Fifthly, before starting the next iteration, the regions are refined. In our approach the region refinement follows from the quadtree decomposition. Details can be found in the next Subsection 2.5.

In our actual implementation we separate in addition the refinement of the regions and the refinement of the Hough table to different iteration steps. Only if no Hough peak can be detected anymore a refinement of the regions is performed.

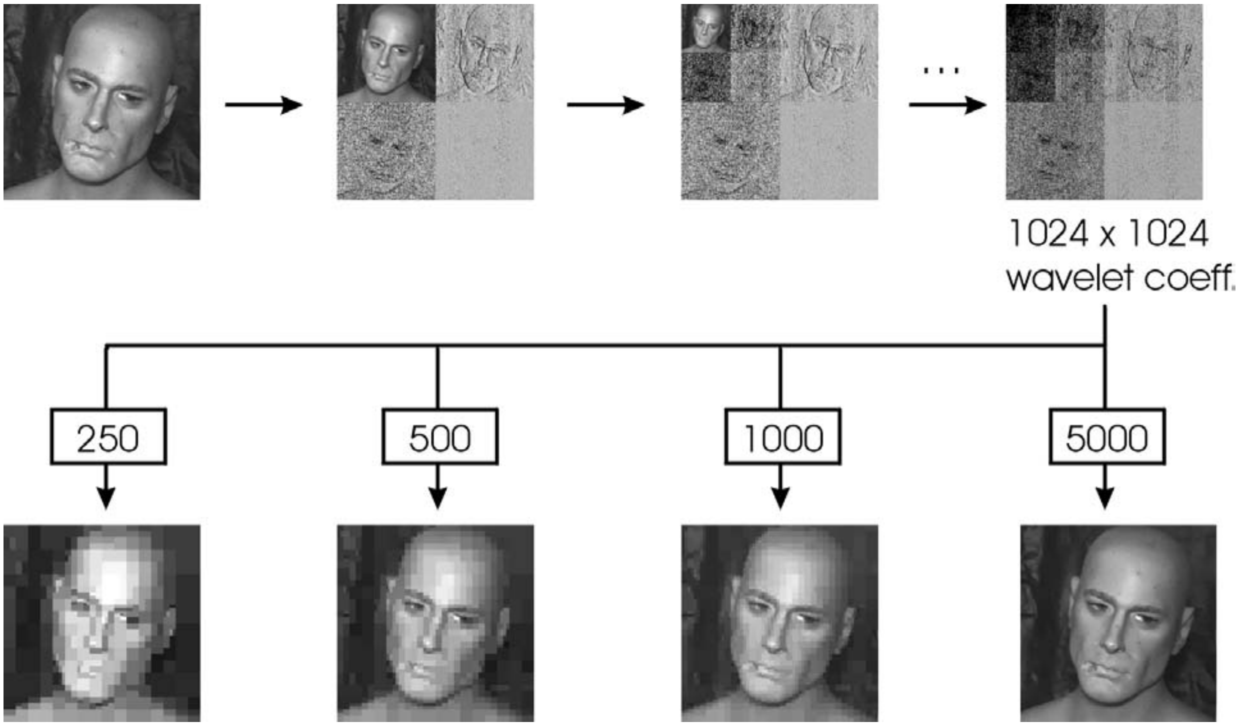

Figure 3: Nonstandard Haar wavelet decomposition and reconstruction from the greatest 250, 500, 1000 and 5000 wavelet coefficients.

2.5 Implementation of Hierarchical Segmentation

The principle of the Haar wavelet based region determination can be best illustrated by Figure 3 for an intensity image: After performing a so called nonstandard wavelet decomposition [18] into 1024×1024 wavelet coefficients we reconstruct the image from the greatest 250, 500, 1000 and 5000 wavelet coefficients resulting in the same amount of regions; the other coefficients are set to zero. The greatest wavelet coefficients either represent big regions or sharp variations in small regions so that regions of different sizes arise. Since the images consisting of more wavelet coefficients also contain the coefficients of the images consisting of less wavelet coefficients, these images subdivide the regions of their lower resolution counterparts. In this way the original intensity image is hierarchically segmented.

In order to make the region determination independent of the viewpoint of the range images, we perform the described wavelet based segmentation on the maximal (in absolute value) curvature maps of the given range images (see Figure 4). The artifacts in the reconstructed images result from a special treatment of invalid data points.

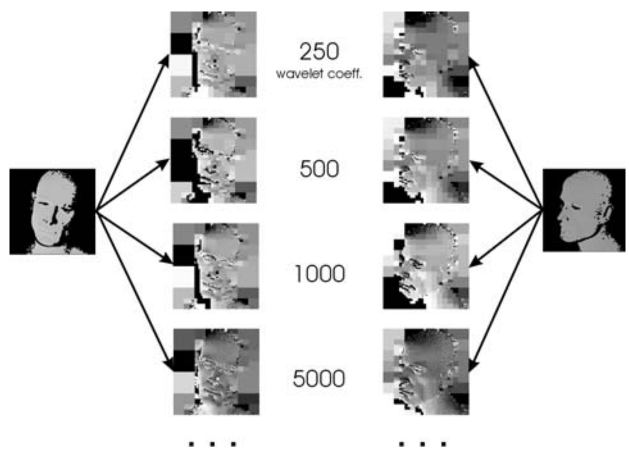

Figure 4: The wavelet based, hierarchical segmentation is independently performed on the maximal curvature maps of two range images from different viewpoints.

Until now we only know that there are some regions in the reconstructed images but we have no easy access to these regions. To get this access we perform a quadtree decomposition [18] of the reconstructed curvature map with 250

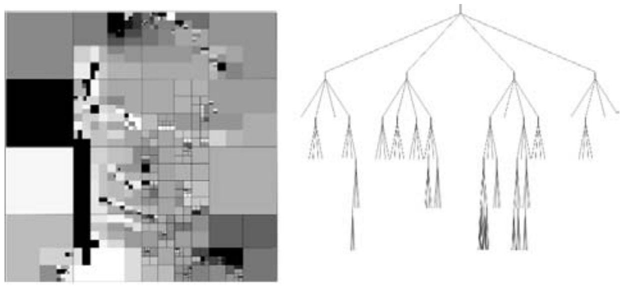

Figure 5: A quadtree representation permits easy access to the regions as the leafs of the quadtree.

wavelet coefficients (see Figure 5): The root of the quadtree corresponds to the whole image; it contains the mean value (here mean maximal curvature) of the entire image. The root has four children, each of them corresponds to one of the image's four quadrants and contains the mean value of the quadrant. Each child has again four children, and so on. The decomposition of a child ends when there is no further structure in the parent's quadrant corresponding to the child. Such a child is called a leaf. Performing the *identical*[1] quadtree decomposition on the map of local frames and on the map of points (range image) we get an elegant access to the local frames and the centers of mass of each region. In addition, we can easily refine the regions by decomposing all leafs in the quadtree decomposition of local frames and of 3D points.

After the quadtree decomposition the regions can be easily transformed into each other and the transformations can be registered into the Hough table (see Figure 6).

2.6 Hough Table Resolution

The number of Hough cells N_H in the Hough table determines the memory consumption (we need at least 2 bytes for the counter in each Hough cell) as well as the resolutions of the transformation dimensions $\alpha, \beta, \gamma, t_x, t_y$ and t_z (there are approximately $\sqrt[6]{N_H}$ intervals in each dimension). We are interested in the minimal number of Hough cells that are necessary so that

[1]The decision for a decomposition of a child is made on base of the curvature map and not on the map of local frames or the map of points itself.

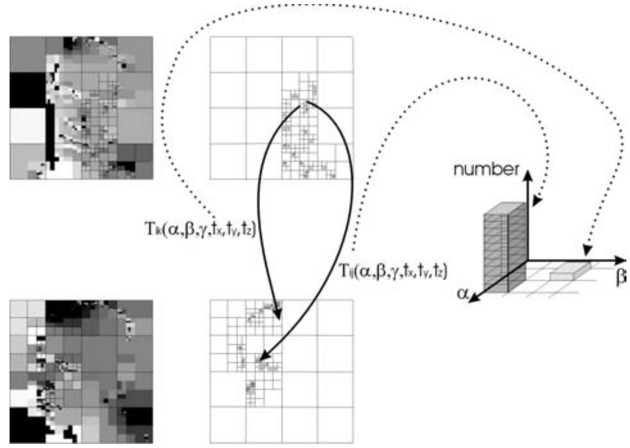

Figure 6: Registration by transforming areas into each other.

the transformations registered in the Hough table will not randomly accumulate to a peak comparable to the expected signal peak. To analyze this problem we assume that the transformations registered in the Hough table are randomly distributed with equal probability p to fall in a certain of the N_H Hough table cells, i.e.

$$p = \frac{1}{N_H}. \tag{4}$$

Thus, if we calculate N_T transformations, the probability that exactly l of them randomly fall in a certain Hough table cell is given by the binomial distribution,

$$p_l = \binom{N_T}{l} p^l (1-p)^{N_T - l} \tag{5}$$

and the probability that k or more than k transformations fall in a certain Hough table cell is,

$$p_{\geq k} = \sum_{l=k}^{N_T} p_l. \tag{6}$$

Since we transform each region of image 1 in each region of image 2, the number of transformations N_T is given by

$$N_T = 2 N_{R_1} N_{R_2} \tag{7}$$

where N_{R_i} is the number of regions in image i and the factor 2 results from the fact that the local frames that are the base of the transformation calculations are not uniquely determined [8].

If we assume that the probability that k or more than k transformations fall in a certain Hough table cell is independent of the same event in another cell[2], the probability that there are m cells in the Hough table with k or more than k transformations is again given by a binomial distribution,

$$\tilde{p}_{\geq m} = \binom{N_H}{m} p_{\geq k} (1 - p_{\geq k})^{N_H - m}. \tag{8}$$

For the expectation value (as well as for the standard deviation) of the binomial distribution there are easy analytical expressions [14]. Therefore, we get for the expected number of Hough cells with k or more than k entries in the Hough table,

$$E_{\geq k} = N_H \, p_{\geq k}. \tag{9}$$

The expectation value $E_{\geq k}$ for k given by the expected signal should be much less than 1 to make sure that there is no random peak in the Hough table with the same size as the expected signal N_S. The expected signal is given by the number of regions in the overlapping area of the two images. Since we do not know this number beforehand we have to prescribe it in actual calculations of N_H.

Putting Eqs. (4)–(9) together we get an equation that cannot be analytically solved for N_H even if we approximate the binomial distribution in Eq. (5) by a Gaussian distribution and the sum in Eq. (6) by an integral. Therefore in our actual calculations we choose N_H in a way that the signal to noise ratio has a prescribed value s. As noise we define the expected value of transformations in a Hough cell plus three times its standard deviation. Therefore we have for the signal to noise ratio:

$$s = \frac{N_S}{\mu_{\text{Underground}} + 3\sigma_{\text{Underground}}} \tag{10}$$

with

$$\mu_{\text{Underground}} = N_T \, p, \tag{11}$$
$$\sigma_{\text{Underground}} = \sqrt{N_T \, p \, (1-p)}. \tag{12}$$

[2] This is certainly not the case, but for $N_T \gg k$ this is a reasonable approximation.

This can be easily solved for $N_H = 1/p$. We give a typical example: for $s = 10, N_{R_1} = N_{R_2} = 10^3, N_S = N_{R_i}/50 = 20$ we get $N_H \approx 6.5 \times 10^6$. This means we have $\sqrt[6]{N_H} \approx 14$ intervals per dimension in the Hough table which corresponds to an angle resolution of approximately $360^\circ/14 = 25^\circ$. Inserting the calculated N_H in Eq. (9) we get $E_{\geq N_S} < 10^{-250}$. Therefore, for a signal to noise ratio of $s = 10$ we can be sure that there is no random peak in the Hough table that can be confused with the signal peak.

3 Conclusions

In this paper, we have presented an iterative hierarchical registration algorithm that has the potential to combine the two steps of usual registration in one approach: the rough estimation of the searched transformation and its precise determination. The algorithm is hierarchical in two ways: the objects to be registered and, independently, the transformation parameter space are hierarchically decomposed. In determining the transformation from a Hough table the algorithm is expected to be very robust against outliers.

The practical feasibility of the whole algorithm has to be proven in the near future.

References

[1] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239 – 256, Feb. 1992.

[2] K. Brunnström and A. J. Stoddart. Genetic algorithms for free-form surface matching. In *13th Int. Conference on Pattern Recognition*, pages D673 – 689, Vienna, Austria, 1996. http://www.ee.surrey.ac.uk/pubvision/-papers/brunnstrom-icpr96.ps.Z.

[3] S. Campagna and H.-P. Seidel. Generating and displaying progressive meshes. In H.-P. Seidel, B. Girod, and H. Niemann, editors, *3D Image Analysis and Synthesis '97*, pages 99 – 108. Infix, Nov. 1997.

[4] Y. Chen and G. Medioni. Object modelling by registration of multiple range images. *Image and Vision Computing*, 10(3):145 – 155, 1992.

[5] Chin Seng Chua and R. Jarvis. Point signatures: a new representation for 3d object recognition. *International Journal of Computer Vision*, 25(1):63 – 85, 1997.

[6] M. P. do Carmo. *Differentialgeometrie von Kurven und Flächen*. Vieweg Studium, 1993.

[7] M. Eck, T. DeRose, T. Duchamps, H. Hoppe, M. Lounsbery, and W. Stuetzle. Multiresolution analysis of arbitrary meshes. In *Proceedings of SIGGRAPH '95*, Annual Conference Series, pages 173 – 182. ACM SIGGRAPH, 1995.

[8] J. Feldmar and N. Ayache. Rigid, affine and locally affine registration of free-form surfaces. *International Journal of Computer Vision*, 18(2):99 – 119, 1996.

[9] W. E. L. Grimson. On the sensitivity of the hough transform for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3):255 – 274, March 1990.

[10] H. Hoppe. Progressive meshes. In A. Glassner, editor, *Proceedings of SIGGRAPH '96*, Annual Conference Series, pages 99 – 108. ACM SIGGRAPH, August 1996.

[11] J. Illingworth and J. Kittler. The adaptive hough transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):689 – 698, 1987.

[12] S. Karbacher and G. Häusler. Fast and simple computation of principal directions on polygon meshes. *Chair for Optics, Annual report*, 1999.

[13] X. Laboureux, S. Seeger, and G. Häusler. Computation of curvatures from 2.5d raster data. *Chair for Optics, Annual report*, 1999.

[14] F. Reif. *Fundamentals of Statistical and Thermal Physics*. McGraw-Hill, 1985.

[15] H. Schönfeld, G. Häusler, and S. Karbacher. Reverse engineering using optical 3d sensors. In *Three-Dimensional Image Capture and Applications*, volume 3313 of *Proceedings of SPIE*. R. N. Ellson and J. H. Nurre, 1998.

[16] G. Stockman, S. Kopstein, and S. Benett. Matching images to models for registration and object detection via clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(3):229 – 241, 1982.

[17] A. J. Stoddart and K. Brunnström. Free-form surface matching using mean field theory. In *British Machine Vision Conference*, pages 33 – 42, Edinburgh, UK, 1996. http://www.ee.surrey.ac.uk/pub/-vision/papers/stoddart-mft-bmvc96.ps.Z.

[18] E. J. Stollnitz, T. D. DeRose, and D. H. Salesin. *Wavelets for Computer Graphics - Theory and Applications*. Morgan Kaufmann Publishers, 1996.

[19] G. Subsol, J. P. Thirion, and N. Ayache. A scheme for automatically building three-dimensional morphometric anatomical atlases: application to a skull atlas. *Medical Image Analysis*, 2(1):37 – 60, 1998.

[20] J. P. Thirion. New feature points based on geometric invariants for 3d image registration. *International Journal of Computer Vision*, 18(2):121 – 137, 1996.

[21] Wu Wang and S. S. Iyengar. Efficient data structures for model-based 3-d object recognition and localization from range images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(10):1035 – 1045, Oct. 1992.

[22] Y.Lamdan and H. J. Wolfson. Geometric hashing: A general and efficient model-based recognition scheme. *2nd Int'l Conf. Comput. Vision*, pages 238 – 249, 1988.

[23] Zhengyou Zhang. Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13(1):119 – 152, 1994.