Registration of 3-D Partial Surface Models Using Luminance and Depth Information

S.Weik
Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Universität Hannover - Germany
email: weik@tnt.uni-hannover.de

Abstract

Textured surface models of three-dimensional objects are gaining importance in computer graphics applications. These models often have to be merged from several overlapping partial models which have to be registered (i.e. the relative transformation between the partial models has to be determined) prior to the merging process. In this paper a method is presented that makes use of both camera-based depth information (e.g. from stereo) and the luminance image. The luminance information is exploited to determine corresponding point sets on the partial surfaces using an optical flow approach. Quaternions are then employed to determine the transformation between the partial models which minimizes the sum of the 3-D Euclidian distances between the corresponding point sets. In order to find corresponding points on the partial surfaces luminance information is linearized. The procedure is iterated until convergence is reached. In contrast to only using depth information, employing luminance speeds up convergence and reduces remaining degrees of freedom (e.g. when registering sphere-like shapes).

Index Terms - motion estimation, image registration, range, intensity, 3D

1. Introduction

For future 3D virtual environments highly realistic three-dimensional surface models of real objects are needed. These *real* models as opposed to synthetic CAD data make virtual scenes more realistic and lifelike. Apart from the mere 3D shape the original image information has to be mapped onto the model's surface (texture mapping) to yield a highly realistic result. The problem in creating 360 degree textured surface models of 3D objects stems from the fact that most of the current systems are not capable of reconstructing the complete model in a single step without some positioning device; instead a sequence of overlapping partial surface models is created, each describing the object from a certain viewpoint. These must be registered to find the correct motion between subsequent viewpoints.

Most of the existing literature dealing with this problem of registering and integrating partial models into complete, closed surface models is restricted to either the use of range data [5][3] or luminance information [11][12][13][14][15]. The first method tries to identify corresponding points on the partial surfaces mostly using iterative closest point algorithms. They operate shape based and thus lack the ability of exactly registering sphere-like shapes. Especially in the case of textured models this leads to inaccuracies in areas where the texture taken from one viewpoint merges with that taken of the neighbouring viewpoint. The second group is mostly based on the essential matrix but often fails to provide the accuracy to integrate the required number of viewpoints into a closed, consistent surface model, when not enough corresponding 2D image points can be found.

Relatively little work has been published in the area of integrating both range and intensity data in the registration process[4][2][1].

A novel approach for registration of textured, partial surfaces to reconstruct an integrated surface representation of three-dimensional objects is presented in this paper. It takes advantage of the ability of some optical range sensors (e.g. a stereo approach) to record luminance and depth information simultaneously. In this research only depth from stereo was used, but the proposed approach should be applicable to laser scanners with combined range and intensity measurement as well.

The organisation of this paper is as follows. Section 2. describes the generation of partial surface models using depth from stereo. In section 3. corresponding point set registration together with the novel approach to select corresponding points making use of the intensity information is explained. Some experiments that have been performed on both synthetic data and real data of a human head are presented in section 4.

0-8186-7943-3/97 $10.00 © 1997 IEEE

2. Partial surface models from stereo

To reconstruct partial surface models a stereo approach is used in this research. It is described in detail in [6]. A calibrated two camera system is used for taking several images of the object to be modeled. Each camera is represented by a pinhole camera model. In fig. 1 the used camera model as proposed by [8] is depicted. It basically consists of the four vectors c, a, h and v. The focal point c describes the camera's position in world coordinates, while the optical axis a gives a description of the viewing direction. The image plane is spanned by u and v. The projection T of a global point P into the image plane calculates as follows:

$$T_h = \frac{(c-P)\cdot h}{(P-c)\cdot a} \tag{1}$$

$$T_v = \frac{(c-P)\cdot v}{(P-c)\cdot a} \tag{2}$$

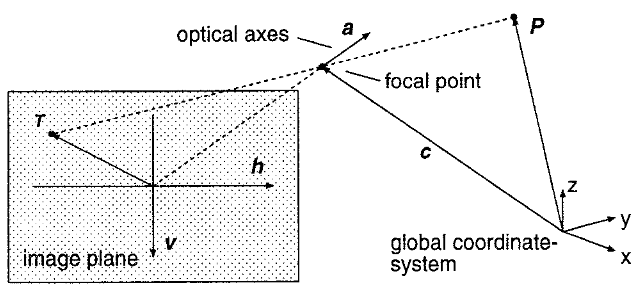

Figure 1. cahv camera model

Two such cameras are rigidly connected to form a stereo sensor as depicted in fig. 2. Prior to taking the images all parameters of the two camera system are calibrated using Tsai-calibration [9]. A global point P is projected into both camera planes along the so-called lines of sight S_l and S_r yielding the image points p_l and p_r. On the other hand, given two corresponding image points p_l and p_r the global surface point P can be calculated since the relative orientation of the left and right camera is known by calibration.

To reconstruct the complete 3D-geometry from the two images all corresponding points have to be found using a blockmatching algorithm. Fortunately this search is constrained to the so called epipolar line which is the projection of the respective line of sight into the other image plane. Using a rectification process both images and the *cahv* parameters are transformed in such a way that the epipolar lines are projected onto horizontal image lines. This further reduces the costs of finding the displacement between corresponding image points.

A dynamic programming algorithm is then employed to perform a line-wise optimisation of all matches on that line.

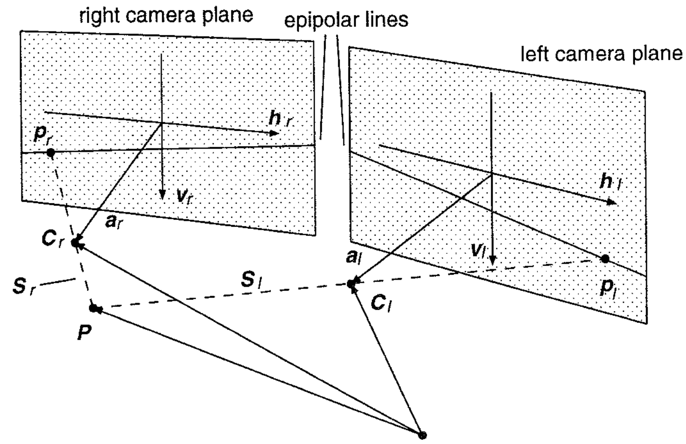

Figure 2. The stereo camera model

A number of different constraints are exploited to enhance the quality of the result [10]. The resulting map of displacement vectors from left to right image (the so-called disparity map) is then compared to the one that was obtained in the opposite direction (i.e. from right to left image). Displacement vectors that cannot be verified in both maps are eliminated to increase measurement reliability. Missing parts are interpolated to obtain a dense map.

Through calibration each corresponding pair of image points can be projected into 3D space yielding a cloud of 3D points in the respective camera coordinate system. Alternatively this cloud of points can be stored in a so called depth map where each pixel describes the distance of the camera's focal point to the surface of the object. Figure 3 shows a left/right image pair taken of a human head. In figure 4 on the left the respective camera oriented depth map is depicted. For the sake of clarity the depth map was transformed into a 3D surface mesh which is shown in a shaded view on the right side. The original image information can later be remapped onto the surface of the model to yield a photorealistic impression (texturing).

Only the left luminance image together with the calculated depth map is used for further processing. Several such partial surfaces, created with the described stereo sensor from different viewpoints, are needed for an entire 3D shape description of an object. The determination of the relative pose within a sequence of such combined range and intensity images is the first processing step in order to fuse the partial models into a complete 360 degree textured surface model.

Figure 3. Left and right luminance image of a human head

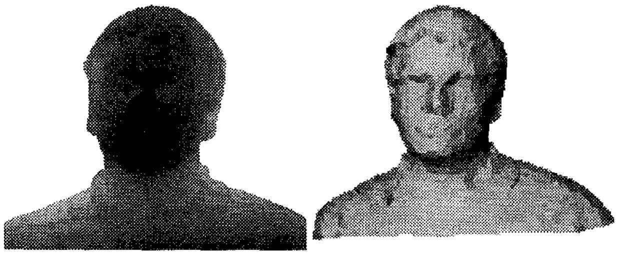

Figure 4. Calculated depth map and subsampled cloud of surface points

3. Registration of 3D partial surface models

In this research the resgistration of 3D partial surface models is divided into two steps. In a first step corresponding points have to be identified on the surface of the two partial models. In the second step the two extracted point sets have to be aligned to each other to minimize a cost function, e.g. the sum over the euclidian point distances. One way of doing this would be to choose closest points, register the point sets and repeat the process until convergence is reached. The very time consuming search for closest points is the main problem with that kind of algorithms. It is therefore replaced by an intensity driven selection of corresponding point sets.

The next section shortly describes how to register corresponding point sets using an quaternion based approach while section 3.2. introduces the newly proposed selection of corresponding point sets.

3.1. Corresponding point set registration

The following quaternion based method to register two corresponding point sets is part of many publications so only a short summary is given here. The complete algorithm can be found e.g. in [5][7].

A quaternion is a four vector $q_R = [q_0 q_1 q_2 q_3]^t$, where $q_0 \geq 0, q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$. It describes a rotation axis and an angle to rotate around that axis. From such a unit rotation quaternion a 3×3 rotation matrix R can be calculated as follows:

$$R = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 - q_0 q_3) & 2(q_1 q_3 + q_0 q_2) \\ 2(q_1 q_2 + q_0 q_3) & q_0^2 + q_2^2 - q_1^2 - q_3^2 & 2(q_2 q_3 + q_0 q_1) \\ 2(q_1 q_3 - q_0 q_2) & 2(q_2 q_3 + q_0 q_1) & q_0^2 + q_3^2 - q_1^2 - q_2^2 \end{bmatrix}$$

Together with a translation vector $q_T = [q_4 q_5 q_6]^t$ the complete registration state vector q is denoted $q = [q_R | q_T]$.

Given a destination point set $D = \{D_i\}$ to be aligned with a corresponding source point set $S = \{S_i\}$ where the numbers of points N_D and N_S in the point sets are equal and the point D_i corresponds with the point S_i for all indices i, the mean square objective function to be minimized is

$$f(q) = \frac{1}{N_S} \sum_{i=1}^{N_S} \|S_i - R(q_R) D_i - q_T\|^2, \tag{3}$$

which is the average point-to-point distance between the corresponding point sets.

The mean values μ_D and μ_S are subtracted from the respective point set. The cross-covariance matrix is calculated of the sets D and S from which the optimal rotation vector $q_R = [q_0 q_1 q_2 q_3]^t$ is derived. The optimal translation vector is calculated as

$$q_T = \mu_S - R(q_R)\mu_D. \tag{4}$$

The calculated registration transformation is applied to the destination point set D to minimze the sum of euclidian distances between the two point sets.

3.2. Luminance based selection of corresponding points

In the following the proposed algorithm to select corresponding point sets using intensity and depth information is described in detail. For each viewpoint to be registered an intensity image together with the calculated depth map is used as a partial surface. The depth map stores for each image point the distance between the camera's focal point and the object surface.

The algorithm tries to register the position of the *destination* camera with respect to the *source* camera which does not change its position. It works as depicted in figure 5:

Given a set $s = \{p_s^{(i)}\}$ of evenly distributed image points $p_s^{(i)}$ in the in the source camera. Each image point $p_s^{(i)}$ is assigned its image intensity $I_s^{(i)}$ and its image gradient $g_s^{(i)}$. From the current source camera position and the respective

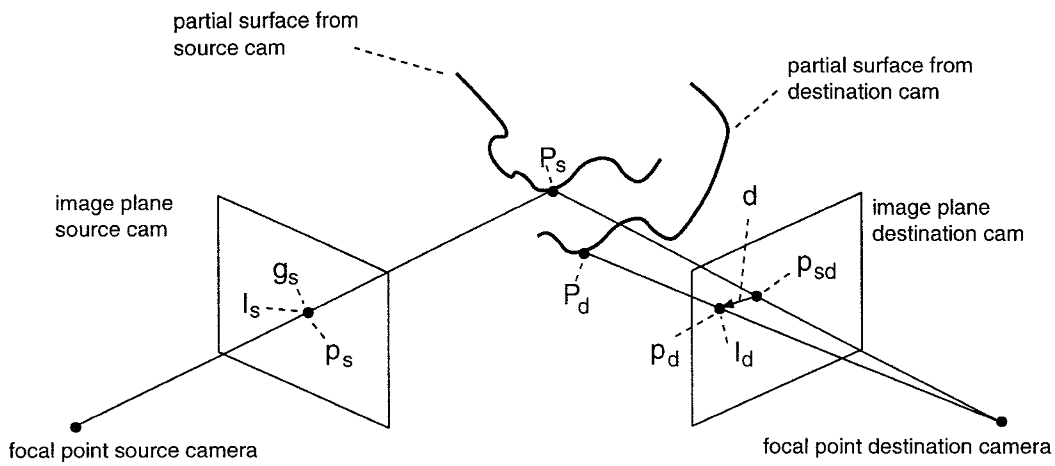

Figure 5. Luminance based selection of corresponding points

depth map a set $S = \{P_s^{(i)}\}$ of surface points $P_s^{(i)}$ is calculated in world coordinates. Each of these points $P_s^{(i)}$ together with its respective image gradient is projected into the destination camera at its current position yielding a set of projected image points $sd = \{p_{sd}^{(i)}\}$ in the destination camera. The projected points $p_{sd}^{(i)}$ generally will have a different image intensity $I_d^{(i)}$ since the destination viewpoint has not been registered completely yet.

The intensity differences

$$\Delta I^{(i)} = I_d^{(i)} - I_s^{(i)} \tag{5}$$

are then used together with the projected image gradient to predict corresponding image points $p_d^{(i)}$:

$$p_d^{(i)} = p_{sd}^{(i)} + d^{(i)}, \tag{6}$$

where $d^{(i)}$ is calculated from

$$\Delta I^{(i)} = g_s^{(i)} \cdot d^{(i)} \tag{7}$$

At the predicted position $p_d^{(i)}$ in the destination image the depth is taken from the destination camera's depth map to calculate a set $D = \{P_d^{(i)}\}$ of destination surface points.

The sum f of Euclidian distances d between the two point sets $S = \{P_s^{(i)}\}$ and $D = \{P_d^{(i)}\}$ is then minimized by a quaternion based approach varying the position of the destination point set by a transformation T . The objective function $f(T)$ to be minimized is:

$$f(T) = \sum_i d(P_s^{(i)}, T(P_d^{(i)})), \tag{8}$$

where T specifies a rigid 3-D tranformation consisting of three translations t_x, t_y, t_z and three rotation angles r_x, r_y, r_z (see section 3.1. on details of T). The minimzation result T is applied to the destination camera and its rigidly connected partial surface. In equation 5 the intensity signal is linearized so the whole process has to be iterated until a convergence criterion is met.

3.3. Selection of surface points with high reliability

In order to make the algorithm robust and efficient mismatches among the corresponding point sets have to be avoided. To meet this demand several measures are taken here.

3.3.1. Selection based on image gradient

To reduce computational costs not all points that are contained in a depth map (approx. 100000) are used for point set matching. Instead a more or less evenly distributed subset of about 3000 points is taken. Thus one important factor concerning the reliability of the algorithm lies in the proper selection of a sample from all available surface points. In this research only points with a high image gradient in the intensity image are used. In addition the selection takes care that the chosen points are more or less evenly distributed over the image. Figure 6 shows an example distribution of selected surface elements. This selection is performed for the source camera only (see fig. 5) the corresponding points in the destination cameras partial surface can be any point of the respective depth map.

3.3.2. Suppression of invisible matches

Making the registration robust with regard to point mismatches requires some extra measures to take. Many problems when determining the relative pose between partial

Figure 6. Distribution of selected surface points

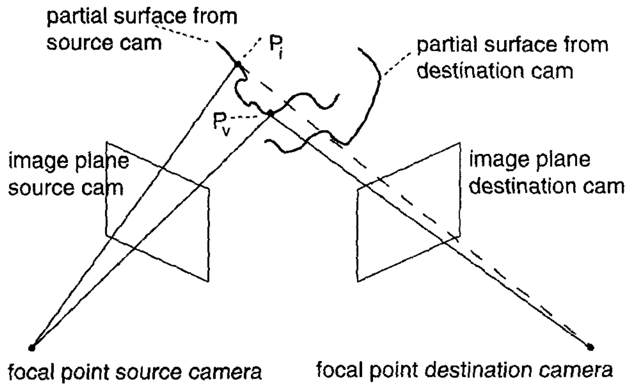

Figure 7. Occluded surface elements during registration

surfaces stem from their overlapping geometry. Not all parts of the surface that are visible from one viewpoint will be seen in the next camera position. On the other hand surface points that are occluded in a camera's viewpoint can not be part of the respective surface model. Only the surface points that are contained in both surface models or that are visible in both cameras respectively are allowed in the corresponding point sets. All other points have to be discarded. Otherwise the attempt to match the partial surfaces would lead to significant errors or even fail.

One way of dealing with that kind of problem is to introduce a threshold for a maximum distance between two corresponding points. Another possibility could be to calculate a histogram of all point to point distances and ignore a certain percentage of point pairs with highest distance. The first solution is only applicable when the partial surfaces are already very close in the beginning. The latter restricts the maximum allowed percentage of a partial surface that is not contained in the other partial model.

The additional intensity information gives an effective way of tackling that problem. In figure 7 a standard situation when registering two surfaces is depicted. The point P_i lies on the source partial surface but can not be seen in the destination camera. Thus no corresponding point can be found on the destination surface model. The point has to be discarded. The point P_v on the other hand has a corresponding element in the other camera and is used for matching.

To identify the points that are occluded in the destination camera the partial model of the source camera is approximated by a triangular mesh. This mesh is rendered into the destination camera at it's current position. A point P_i then can be identified as occluded, since the respective line of sight intersects the partial surface in another point, which is closer to the destination camera.

4. Experimental results

The algorithm has been tested both for a synthetic box-like object and real data of a human face.

4.1. Synthetic data

As synthetic data a box-like object has been used. A synthetic texture has been mapped to the surface. The resulting luminance and depth image is shown in figure 8. A 5 degree rotated view has been matched to the depicted viewpoint.

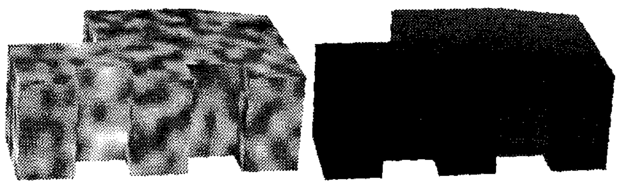

Figure 8. Luminance and depth image of synthetic box

The algorithm converged after about 15 iterations and the rotational deviation could be determined to within 0.1 degree. Figure 9 shows the convergence behavior of the angles for the described example, figure 10 depicts the convergence of the average point to point distance in mm.

4.2. Real data of a human face

To test the algorithm with real data a sequence of a human head rotating in front of a calibrated stereo camera has been used. The corresponding depth map was calculated with the described stereo sensor. In figure 11 two intensity

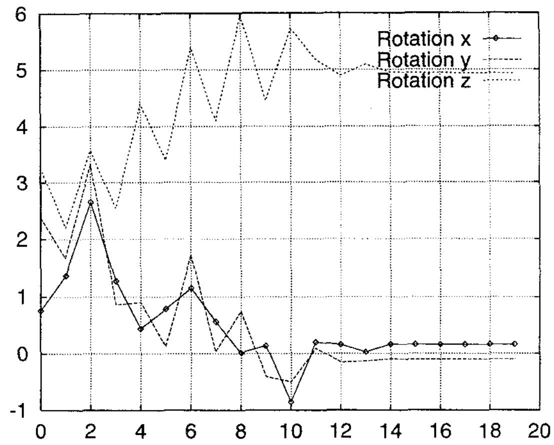

Figure 9. Convergence behavior for synthetic data (angles in degree over iterations)

images for two different viewpoints that have been registered are shown. Below the intensity images the respective depth information is depicted. For the sake of clarity it has been transformed into a triangle mesh and is looked at from the side. It has been printed in a shaded view to show more clearly the 3D character of the partial surfaces.

Between the two viewpoints the stereo sensor was rotated by about 30 degrees. It can be seen from the depth data that each viewpoint covers approximately 30 percent of the head. The covered area rotates together with the viewpoint. Since the images were taken while the person was rotating on a swivel-chair the exact relative movement between the two viewpoints is unknown. Therefore some overlied images of the matched partial surfaces are shown in figures 12

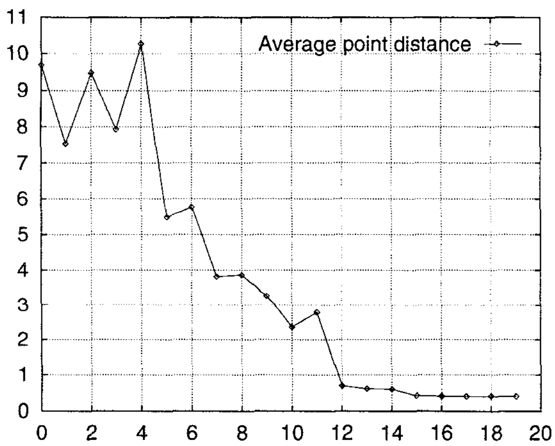

Figure 10. Convergence of average point distance (mm over iterations)

and 13 instead of exact numbers. The dark areas belong to the registered surface.

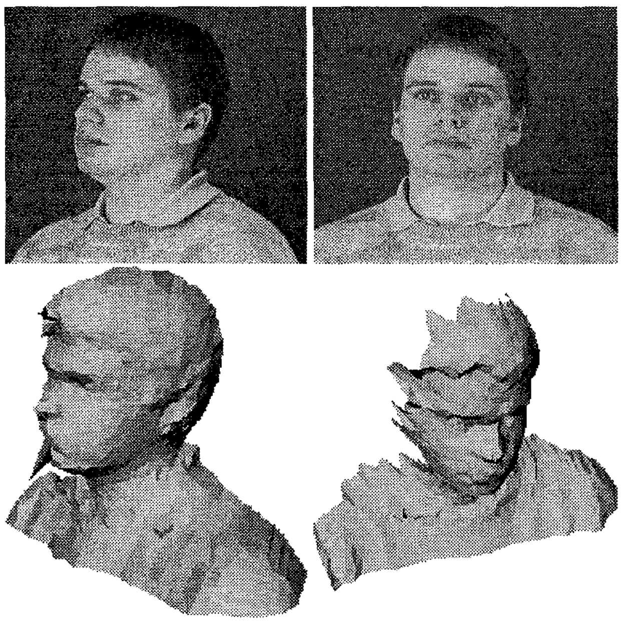

Figure 11. Luminance images and respective triangulated depth maps for the registered viewpoints

5. Conclusion

A novel algorithm for registration of partial 3-D surface models has been presented. It makes use of both luminance and depth information. For a set of highly reliable luminance points on one partial surface the corresponding points on the other partial surface are predicted using the luminance gradient information. The second partial surface are then transformed to minimize the Euclidian sum of distances between the two point sets. This is performed with a quaternion approach. Since intensity information is linearized by its gradient the process has to be iterated until a global convergence criterion is met.

It is neither necessary to calculate corresponding features in the luminance images nor to search for closest points in the depth information. Compared to the time consuming iterative determination of closest points in depth data alone, the straightforward way of choosing corresponding candidates using the image gradient makes the algorithm very fast. Knowing which 3D surface point belongs to a certain image point allows to effectively suppress wrong matches among the point sets. The additional intensity information can be especially useful when partial surfaces of sphere-like

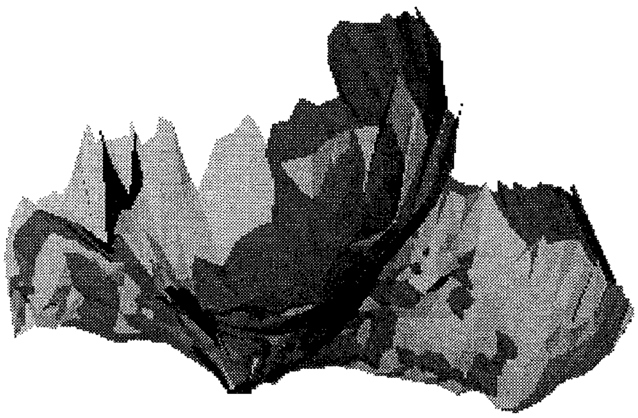

Figure 12. Matched partial surfaces (black, grey). Top view

objects are registered since using depth data alone leaves one or two degrees of freedom.

Important for the reliability of the algorithm is the quality of the intensity images considering the average luminance. The selection of corresponding point sets relies on the image gradient in combination with an intensity difference between subsequent images. Differences in the global intensity of the images lead to mismatches among the corresponding points. When taking the pictures one should take care that each image is given the same exposure. The camera's automatic exposure should be avoided.

In tests with synthetic and real data the algorithm managed to register rotational angles of up to 30 degrees without

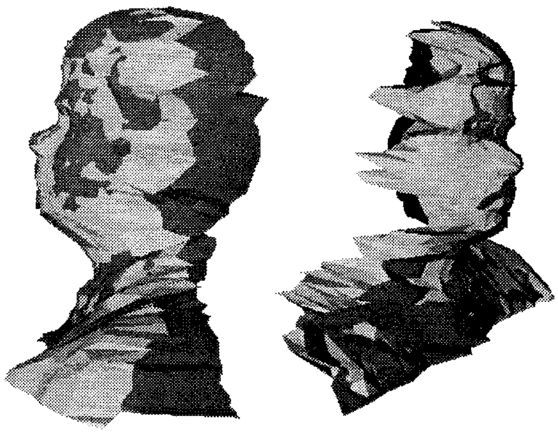

Figure 13. Matched partial models from left and right

assumptions of the new position. It usually converges in 10 to 20 iterations.

6. Future directions

Although the approach performed quite efficient on synthetic data of a box like object and real data of a human head, a lot of further testing is still needed. In both cases the depth data was yielded by a stereo approach. Additionally it should be tested with data sets from a laser scanner that is capable of simultaneously measuring a combined set of intensity and range data.

Furthermore the integration of already registered partial surfaces into a complete database to yield a consistent 360 degree surface model is an important task. Currently only two view experiments could be performed.

In some situations the algorithm tends to stick to local minima. This problem can be reduced through a hierarchical approach. The intensity image is then subsampled and low-pass filtered to reduce high frequencies in the information. After a rough estimation using this subsampled intensity image the original high-resolution image is taken to perform the rest of the registration.

7. Acknowledgments

The author wishes to thank Reinhard Koch and Lutz Falkenhagen for kindly providing the stereo software package and Olaf Hilgenfeld for contributing his head.

References

[1] R. Koch: *Dynamic 3-D scene analysis using synthesis feedback control*, IEEE PAMI, Vol. 15(6), pp. 556-68, June 1993

[2] M.J. Magee, B.A. Boyter, C.H. Chien, J.K. Aggarawal: *Experiments in intensity guided range sensing recognition of three-dimensional objects*, IEEE PAMI, Vol. PAMI-7, No. 6, 1985

[3] G. Blais, M.D. Levine: *Registering multiview range data to create 3D computer objects*, IEEE PAMI, Vol. 17, No. 8, August 1995

[4] G. Godin, M. Rioux, R. Baribeau: *Three-dimensional registration using range and intensity information*, Proceedings of the SPIE, Vol. 2350, pp. 279-90, 1994

[5] P.J. Besl, N.D. McKay: *A method for Registration of 3-D shapes*, IEEE PAMI, Vol. 14, No. 2, February 1992

[6] R.Koch: *3-D Surface Reconstruction from Stereoscopic Image Sequences*, International Conference of Computer Vision ICCV '95 Cambridge, MA., USA, June 1995

[7] B.K.P. Horn: *Closed-form solution of absolute orientation using unit quaternions*, Journal of the Optical Society of America, vol.4 no. 4, pp. 629-642, April 1987

[8] Y. Yakimovski, R. Cunningham: *A system for extracting 3-D measurements from a stereo pair of TV cameras*, Computer Graphics and Image Processing Vol 7, pp. 195-210, 1978

[9] R. Y. Tsai: *A versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses*, IEEE Journal of Robotics and Automation, Vol. RA-3, No. 4, Aug. 1987, pp 323-344

[10] L. Falkenhagen: *Depth Estimation from Stereoscopic Image Pairs Assuming Piecewise Continuos Surfaces*, Paker, Y. and Wilbur, S. (Ed.), Image Processing for Broadcast and Video Production, Hamburg 1994, pp. 115-127, Springer series on Workshops in Computing, ISBN 3-540-19947-0, Springer Great Britain, 1994.

[11] O.D. Faugeras, Q.T. Luong and S.J. Maybank: *Camera self-calibration: Theory and Experiments* in Computer Vision-ECCV '92 (Second European Conference on Computer Vision, Santa Margherita Ligure, Italy, May 19-22, 1992), Springer, Berlin, 1992, 321-334.

[12] P. Beardsley, D. Murray, and A. Zisserman: *Camera calibration using multiple images* in: G. Sandini, ed., Computer Vision-ECCV '92 (SecondEuropean Conference on Computer Vision, Santa Margherita Ligure, Italy, May 19-22, 1992), Springer, Berlin, 1992, 312-320.

[13] Z.Y. Zhang, R. Deriche, O. Faugeras, and Q.T. Luong: *A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry* Artificial Intelligence, vol. 78, no. 1-2, 1995, 87-119.

[14] Q.-T. Luong and O.D. Faugeras: *Self-Calibration of a Camera Using Multiple Images* Proceedings 11th IAPR International Conference on Pattern Recognition, Volume I, Conference A: Computer Vision and Applications (ICPR11, The Hague, The Netherlands, August 30 - September 3, 1992), IEEE Computer Society Press, Los Alamitos, California, 1992, 9-12.

[15] R. Hartley: *Calibration of cameras using the essential matrix* Proceedings, [DARPA] Image Understanding Workshop, San Diego, CA, January 26-29, 1992 (Morgan Kaufmann, San Mateo, CA), 911-915.