A New Technique for Registration and Integration of Partial 3D Models

Soon-Yong Park and Murali Subbarao

Department of Electrical and Computer Engineering
State University of New York at Stony Brook
Stony Brook, NY 11794-2350
E-mail:{parksy/murali}@ece.sunysb.edu

Abstract

A new technique is introduced for registration and integration of multiple partial 3D models of an object. The technique exploits the epipolar constraint for the multiple-view geometry. Partial 3D shapes of an object from multiple viewing directions are obtained using a digital vision system based on parallel-axis stereo. The vision system is calibrated to obtain an initial tranformation matrix for both the stereo imaging geometry and the multiple-view geometry. A multi-resolution stereo matching approach is used for partial 3D shape recovery. The partial 3D shapes are registered approximately using the initial transformation matrix. The initial transformation matrix is then refined by iteratively minimizing the registration error. At this step, a modified Iterative Closest Point (ICP) algorithm is used for matching corresponding points in two different views. A given point in one view is projected to another view using the transformation matrix and a search is made for a closest point in the other view that lies on the epipolar line. A similar idea is used during partial model integration step to obtain improved results. Partial models are represented as linked lists of segments and integrated segment by segment. Experimental results are presented to show the effectiveness of the new technique.

1. INTRODUCTION

Three-dimensional (3D) model reconstruction of objects is a topic of much interest in Computer Vision and Computer Graphics areas. A number of researchers have investigated this topic[1]. In order to construct a complete (i.e. 360 degree view) 3D model of an object, two important problems need to be solved–*registration* and *integration*–of multiple partial 3D models.

For a given direction of view, only the front part of an object is visible; back part and possibly the left and right sides are not visible. Therefore, for a given view, the shape and image data measured corresponds to a partial 3D model of the object. Either the object or the sensor has to be moved to view the back part and the sides of the object. First, the partial 3D models obtained from different coordinate systems will need to be *registered* with respect to a common coordinate system. It involves finding the transformaion parameters for transforming the different reference coordinate systems to the common coordinate system. After registration, the partial models will need to be *integrated* together to obtain a single complete 3D model.

Many researchers have investigated the registration problem. A popular method of registration is the Iterative Closest Point(ICP) method[2]. It finds the transformation parameters that minimize a measure of registration error between corresponding points between two views. This involves solving the *correspondence problem*, i.e., given shape data points in one view, we need to find the shape data points in the other view which correspond to the same physical points on the object. ICP establishes correspondence by finding the closest point in 3D space in one view for a given point in another view. The transformation parameters are refined iteratively to minimize the registration error.

Finding correspondences between two or more shapes is not an easy problem in computer vision. In particular, when the baseline of multi-views is very wide, it is a difficult research topic[3]. There are other approaches to register multiple range image using feature points on shapes[4,5]. They find corresponding points between two shapes which have a close tangent on surfaces. Weik[6] finds correspondences between two partial shapes using luminance information. Our approach is similar to ICP, but it exploits the epipolar geometry between two views to reduce errors and computation in searching for corresponding points. We iteratively search for a point that lies on the epipolar line in one view and closest in some sense to a given point in another view. For this reason, we call our algorithm–*Epipolar Closest Point* (ECP) algorithm.

In this paper, we present a digital vision system for 3D shape reconstruction. The system acquires partial 3D models using stereo imaging, registers them, and integrates them. Initially, we calibrate the vision system for the transformation matrix between different views using Tsai's calibration technique[7]. The result of this calibration step is used as the initial solution for the transformation matrix which is later refined iteratively by the ECP algorithm. The registered partial 3D models are represented as linked lists of segments and integrated slice by slice to obtain a complete surface model. The integration step also removes some erroneous segments in partial 3D models due to image mismatching during stereo correspondence.

2. PARTIAL SHAPES ACQUISITION

Our digital vision system consists of a high-resolution digital still camera, a translation stage to implement a parallel stereo geometry, a rotation stage to change the view of an object, and a slide projector to introduce contrast on the object surface. All components are computer controllable. Figure 1 shows our vision system named SVIS-II.

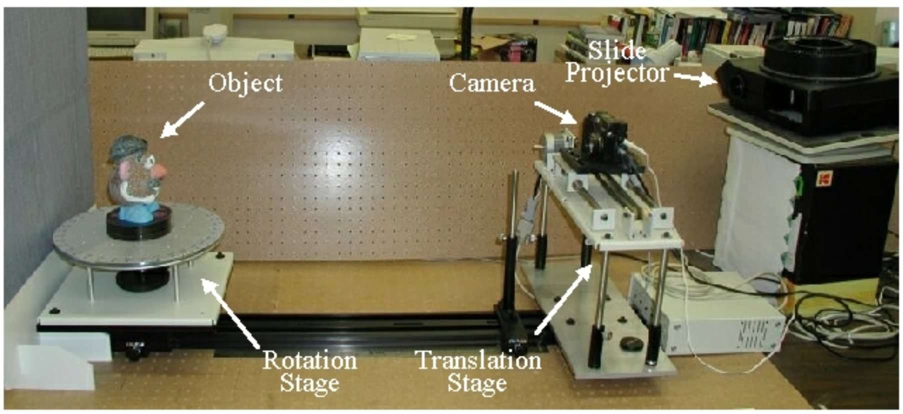


Figure 1: SVIS-II vision system

For partial shape acquisions, we use a multi-resolution stereo matching technique using a Gaussian pyramid[8]. Multi-resolution approach in stereo matching can reduce some mismatching errors and restrict the errors within a pre-defined range. We take a stereo pair from each direction of view, from four directions at 90 degree intervals. The size of an original image is 1280×960 and the size of a range image after stereo matching is 320×240.

Parallel stereo geometry is used to acquire partial shapes as shown in Figure 2. The baseline B of the stereo camera is 76mm and the focal length f of the camera is set to 19.35mm at 'Tele' mode of the zoom lens of the camera. The focal length was calibrated using Tsai's non-coplanar camera calibration technique.

3. REGISTRATION

3.1. Initial registration

We acquire stereo pictures of an object from 4 directions of view by rotating the object by 0, 90, 180, and 270 degrees. However, we model this data acquisition by assuming the object to be stationary but the camera system being rotated around the rotation axis by 0, 90, 180, and 270 degrees respectively. For each of the four rotation positions of the camera system, we associate a coordinate system V_i for $i = 0, 1, 2, 3$. Stereo image analysis is used to compute the partial 3D shapes of the object with respect to V_i. In our experiments, V_i was the same as the right camera coordinate system in the stereo camera.

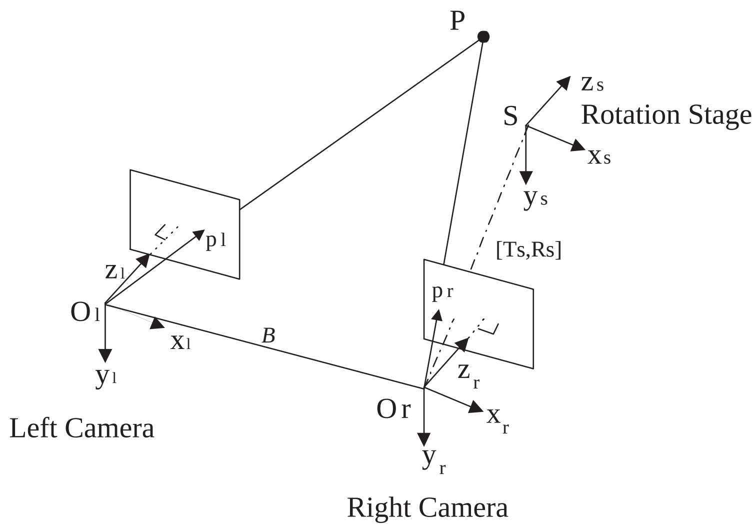


Figure 2: Vision system geometry

The four different partial 3D shapes computed with respect to V_i for $i = 0, 1, 2, 3$ will have to be registered with respect to a common coordinate system. An accurate knowledge of the position and orientation of the rotation axis with respect to V_0 is needed for registering the partial 3D shapes. We assume the orientation of the rotation axis to be almost parallel to the y-axis of V_0 and that the rotation angles 0, 90, 180, and 270 are accurate. The position of the rotation axis is defined by a translation vector T_s. An initial estimate of T_s is obtained through stereo camera calibration using Tsai's technique[7]. This estimate is refined iteratively during the registration step as described later.

Let $\mathbf{P}_{ii}$ denote a point given with respect to V_i and $\mathbf{P}_{ij}$ denote the same point with respect to V_j obtained by registering $\mathbf{P}_{ii}$ to V_j. Also let R_{jk} represent the rotation transformation matrix from V_j to V_k. Now, in order to register a point $\mathbf{P}_{ij}$ to the common coordinate system V_0 to obtain $\mathbf{P}_{i0}$, we use the relation,

$$\mathbf{P}_{i0} = R_{j0}(\mathbf{P}_{ij} - T_s) + T_s \tag{1}$$

where, T_s is the translation vector defined earlier.

3.2. Partial shape representation

The partial 3D shapes are initially registered using an accurate knowledge of R_{j0} but approximate value of T_s from calibration. The registered values are represented in an object centered coordinate system with its y-axis along the rotation axis. The object workspace is assumed to be a 300mm$\times$300mm$\times$300mm cube. We consider the intersection of the partial 3D shapes with 300 horizontal planes 1mm apart and parallel to the $x - z$ plane. The intersections are horizontal contours where each contour consists of a sequence of points. These points are represented using 2D coordinates on the horizontal slice containing them, where each slice corresponds to a constant vertical (y) coordinate in 3D space.

We make linked lists of these points and sort them with respect to x and z coordinates. The sorting is done because we want the linked list to represent points in order on a continous contour on the object's surface. For each slice plane, we generate one linked list for each direction of view, thereby generating 4 linked lists. Let $\mathcal{L}V_i$ be the four linked lists of points for $i = 0, 1, 2, 3$. Integration of these linked lists is merging the lists in such a way that the resulting list represents a closed contour that is an accurate cross-section of the object.

There will be some erroneous points on the list representing short, irregular, false contours that do not belongs to the object. Also, some points on the object's surface could be missing due to gaps in contours. These erros are introduced by stereo mismatching due to occlusion, low contrast, or noise. Therefore, we segment the linked lists to detect and reduce these errors. This is important because refining the registration matrix should be done using only the correct object points.

Initially we have only one linked list of points $\mathcal{L}V_i$ for the ith view. We split the linked lists $\mathcal{L}V_i$ into multiple sublists $\mathcal{L}V_{ij}$ where each sublist represents one connecting contour segment. This splitting is done based on thresholding the distance between two consecutive points on a list. If the distance is larger than the threshold, we divide the linked list into two separate lists under the assumption that the object surface is locally linear and continuous. As a result, a list $\mathcal{L}V_i$ for ith view will be split into multiple lists $\mathcal{L}V_{ij}$ for $j = 0,1, \ldots\ m$ where m is the number of segments on the list. We also generate a new linked list of segments $\mathcal{L}S_i$ pointing all linked lists $\mathcal{L}V_{ij}$ for easy access. After generating linked lists of segments for all views, we remove a linked list if its length is too short, as it is very likely to be due to errors in stereo mismatching.

3.3. Epipolar geometry of views

To refine the registration matrix between two directions of view, we introduce a new technique called ECP(Epipolar Closest Point) method. Our technique exploits the epipolar geometry to reduce errors and computation. Given two segment lists, we compare spatial information of two segments in the image space as well as in the object space. Comparing in the image space is based on the reprojection of one segment to the other's image space. To find out the geometrical information of two segments in the image space, we have to transform the coordiante of one segment to the other's coordinate system[9].

Consider two segment lists $\mathcal{LS}_0$ and $\mathcal{LS}_1$ from V_0 and V_1. Each list consists of many sublists, each sublist representing a contour segment. Therefore we can write

$$\begin{aligned} \mathcal{LS}_0 &= \{\mathcal{LV}_{0j} \mid j = 0, \cdots, m\} \\ \mathcal{LS}_1 &= \{\mathcal{LV}_{1k} \mid k = 0, \cdots, n\} \end{aligned}$$

,where m and n are the number of sublists and $\mathcal{LV}_{ij}$ is the jth sublist of V_i. Let $\mathbf{P}_{00}$ be a point in a list of $\mathcal{LS}_0$, the projection of this point onto the image plane of V_0 be $\mathbf{p}_{00}$, and another projection onto the image plane of V_1 be $\mathbf{p}_{01}$.

Figure 3 shows the geometry of V_0 and V_1. The three points– origin of V_0, origin of V_1, and the point $\mathbf{P}_{00}$– together determine the epipolar plane.

A point $\mathbf{P}_{00}$ in V_0 is denoted by $\mathbf{P}_{0i}$ in V_i. They are related by

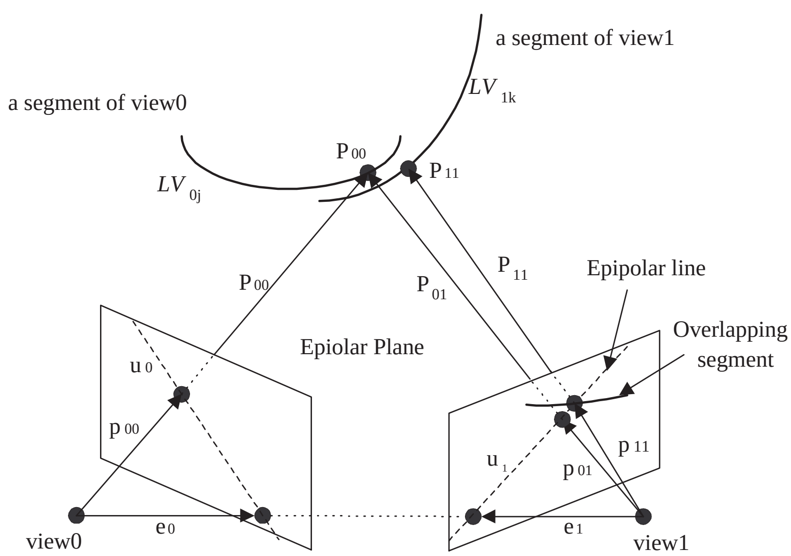


Figure 3: Epiplar geometry of V_0 and V_1

$$\mathbf{P}_{0i} = R_{0i}(\mathbf{P}_{00} - T_s) + T_s. \tag{2}$$

The projection of $\mathbf{P}_{0i}$ onto the image plane in V_i is given by the perspective transformation relation:

$$\mathbf{p}_{0i} = \left(f\frac{\mathbf{P}_{x0i}}{\mathbf{P}_{z0i}}, f\frac{\mathbf{P}_{y0i}}{\mathbf{P}_{z0i}}, f \right), \tag{3}$$

Setting $i = 1$ in the above equations, we can get the epipolar line $\mathbf{u}_1$ on the image plane in V_1 by transforming the image vector $\mathbf{p}_{01}$ as

$$\mathbf{u}_1 = E\mathbf{p}_{01} \tag{4}$$

where, E is a essential matrix and it is represented as

$$E = R_{01}S, \tag{5}$$

and

$$S = \begin{pmatrix} 0 & -T_{sz} & T_{sy} \\ T_{sz} & 0 & -T_{sx} \\ -T_{sy} & T_{sx} & 0 \end{pmatrix}.$$

If two points $\mathbf{P}_{00}$ and $\mathbf{P}_{11}$ correspond to the same physical point on the object, the projections of the points onto the same image plane should have the same coordinate. Therefore $\mathbf{p}_{01}$ and $\mathbf{p}_{11}$ should be the same vectors.

However, because of registration error due to errors in initial calibration caused by stereo mismatching, distortion of the camera lens, and noise, there will be an error between two vectors. To minimize the registration error, we introduce an iterative minimization technique based on the epipolar geometry of the two directions of view. In Figure 3, we can expect the vector $\mathbf{p}_{11}$ to be on the epipolar line $\mathbf{u}_1$ if two vectors are different representations of the same point, because the epipolar line is the transformation of the vector $\mathbf{p}_{01}$ to the V_1 image plane.

3.4. Refining registration matrix

Earlier, we briefly mentioned the calibration of our vision system. In our vision system, we assume that the rotation axis is almost parallel to the y-axis of the camera coordiante system and the interval of each rotation is exactly 90 degree. Only the translation T_s from the origin of the common camera coordinate system to the center of rotation can have an error.

To refine the translation matrix, we search for two closely overlapping segments from segment lists of two different views and compute a translation error between them. This error T_ϵ is computed for all possible overlapping segments from adjacent pairs of views and averaged. This error is assumed to have a Gaussian distribution. The average error is iteratively minimized until it converges close to zero.

Consider segment lists in two views V_0 and V_1. As before, consider a point $\mathbf{P}_{00}$ in V_0 which is transformed to V_1 to get $\mathbf{P}_{00}$ and then projected onto the image plane in V_1 to get $\mathbf{p}_{01}$ and the corresponding epipolar line $\mathbf{u}_1$. Given a segment in $\mathcal{L}V_{0j}$ that contains $\mathbf{P}_{01}$ in V_0, we find a closely overlapping segment $\mathcal{L}V_{1k}$ in V_1 as shown in Figure 4. We fit a line to the segment $\mathcal{L}V_{0j}$, projected the line onto the image plane, and find the intersection of the projected line with the epipolar line at point $\mathbf{p}_{11}$. This point $\mathbf{p}_{11}$ can be back projected onto the segment $\mathcal{L}V_{1k}$ to determine the point $\mathbf{P}_{11}$ in 3D space.

If we assume that two vectors $\mathbf{P}_{00}$ and $\mathbf{P}_{11}$ in 3D space correspond to the same physical point on the obejct, we can consider the difference of the two vectors as a translation error between two view coordinates. Let $T_\epsilon|_{01}$ be a translation error between two views, then

$$T_\epsilon|_{01} = \mathbf{P_{01}} - \mathbf{P_{11}} \tag{6}$$

Now, we are going to find the translation matrix T_{01} between two object points from $\mathbf{P}_{00}$ to $\mathbf{P}_{11}$. Let $T_s{}'$ be a refined translation matrix between the first view coordiante system V_0 and the rotation center.

$$T_s = T_s{}' + T_\epsilon \tag{7}$$

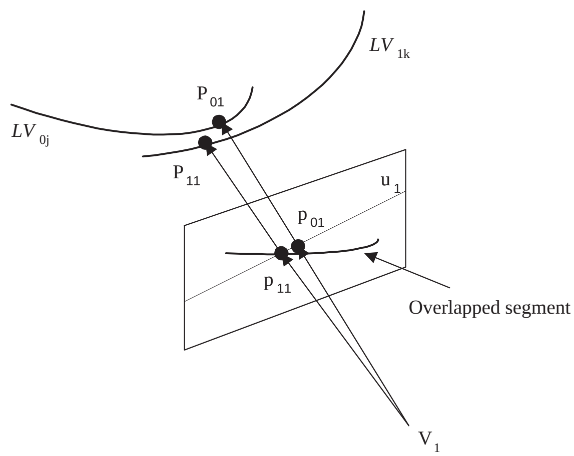


Figure 4: Image plane of view 1

,where T_s is the translation matrix which is calibrated by Tsai's calibration method and T_ϵ is a translation error in T_s. To find the translation between two camera coordinate system, we use the transformation between two vectors

$$\begin{aligned}\mathbf{P_{11}} &= R_{01}(\mathbf{P_{00}} - T_s) + T_s \\ &= R_{01}\mathbf{P_{00}} + (I - R_{01})T_s \\ &= R_{01}\mathbf{P_{00}} + T_{01}.\end{aligned} \tag{8}$$

, where T_{01} is the translation from the vector $\mathbf{P}_{00}$ to $\mathbf{P}_{11}$ and can be expressed as

$$\begin{aligned}T_{01} &= T_{01}' + T_\epsilon|_{01} \\ &and \\ T_\epsilon|_{01} &= (I - R_{01})T_\epsilon \\ T_\epsilon &= (I - {R_{01}}^T)\, T_\epsilon|_{01}\,.\end{aligned} \tag{9}$$

Where, ${R_{01}}^T$ is the transpose of R_{01} and $T_\epsilon|_{01}$ is a translation error in the matrix T_{01}.

In Equation 9, we find the matrix T_ϵ from the epipolar geometry of the overlapping segments, and compute the translation matrix T_s by equation 7. After getting the matrix T_s, we re-register all partial shape to the common coordinate system based on the new translation matrix. The matrix is computed again iteratively until the error T_ϵ converges close to zero. An algorithm for refining the translation matrix is summarized below.

do
  for j = 0 to m : number of segments on $\mathcal{LS}_0$
  for each point on a segment $\mathcal{LV}_{0j}$
    for k = 0 to n : number of segments on $\mathcal{LS}_1$
    for each point on a segment $\mathcal{LV}_{1k}$
      if Overlapping segment($\mathbf{p_{01}}$,$\mathcal{LV}_{1k}$)
        find intersecting point $\mathbf{p_{11}}$
        $T_\epsilon|_{01} \leftarrow \mathbf{P_{01}} - \mathbf{P_{11}}$
        $T_\epsilon \leftarrow (I - {R_{01}}^T)\, T_\epsilon|_{01}$
  $\hat{T}_\epsilon \leftarrow$ averaged T_ϵ
  ${T_s}^{(t+1)} \leftarrow {T_s}^{(t)} + \hat{T}_\epsilon$:update $(t+1)$th iteration
  Register again all segment lists of V_0 and V_1

while ($\hat{T}_\epsilon$ not close to zero).

4. INTEGRATION

4.1. Partial shape integration

After finding the T_s matrix that minimizes the registration error for all views, we integrate partial shapes to obtain a complete 3D shape. The basic ideas used in this step are similar to those in the registration step. Integration is also done slice by slice by merging partial shapes of views V_0 and V_1 first, V_1 and V_2 next, then V_2 and V_3, and finally V_3 and V_0. Consider the integration of partial shapes of views V_0 and V_1. we have two linked lists $\mathcal{L}V_{0i}$ and $\mathcal{L}V_{1j}$, where i and j are the index of segments for each view. For a point $\mathbf{P}$ in the linked list $\mathcal{L}V_{0i}$, we find points on the $\mathcal{L}V_{1j}$, which are close to the point $\mathbf{P}$. As before, let $\mathbf{P}_{00}$ be a point vector in V_0 which becomes $\mathbf{P}_{01}$ after transforming to V_1, and let $\mathbf{p}_{01}$ be its projection onto the image plane in V_1.

Finding points close to $\mathbf{p}_{01}$ is similar to the method in the registration step. It is based on two criteria. We find the epipolar line $\mathbf{u}_1$ corresponding to $\mathbf{p}_{01}$, and find a segment from the list of second view that intersects the epipolar line close to $\mathbf{p}_{01}$. As the first criterion, we use $\mathbf{c}1$ given by a scalar product of two vectors

$$\mathbf{c}1 = \mathbf{u}_1 \cdot \mathbf{p}_{11}. \tag{10}$$

We find a set of close points $\mathbf{p}_{11}$ for which $\mathbf{c}_1$ is smaller than a pre-defined threshold. As the second criterion, we use $\mathbf{c}_2$ which is the Euclidean distance between two points in 3D space,

$$\mathbf{c}2 = dist(\mathbf{P}_{01}, \mathbf{P}_{11}). \tag{11}$$

The distance has to be less than a predefined threshold. Both criteria have to be satisfied for merging the two segment lists. If they are satisfied, the linked list $\mathcal{L}V_{0i}$ is connected to the list $\mathcal{L}V_{1j}$. The connecting point is the closest point on the $\mathcal{L}V_{1j}$ segment from a point on the $\mathcal{L}V_{0i}$ segment.

We compare all linked lists bewteen two views $\mathcal{L}V_{0i}$ and $\mathcal{L}V_{0j}$ for $i = 0, ..., m$ and $j = 0, ..., n$ one by one to find possible connections. After connecting the linked lists of V_0 and V_1, we check the next segment lists between V_1 and V_2, V_2 and V_3, and finally V_3 and V_0 on the same slice. After the last connection, we check the diatance between the staring point of V_0 and the ending point of V_3. If it is small, they are connected to close the segment lists. This results in a closed contour representing the horizontal cross-section of the object. After connecting all segment lists together, we remove again some short-length segments. A Gaussian low pass filter is applied to the connected contours to get a smoothed 3D shape after integration.

4.2. Mesh generation

Surface meshes of shape are generated between two adjacent slices contour by contour, and from top to bottom. Given two slices which are facing each other, we pick two contours from both slices. At the beginning point of one contour, we search for the closest point and the next one from the other and make a triangle using three points. All points are registered as vertices for a mesh of triangles.

5. TEST RESULTS

We put a 'Head' object on the rotation stage for our system for 3D model acquisition. After taking 4 stereo pairs from 4 directions of view, we separate the object area from the background. For separating the object area, we use the 'Depth from Focus' technique. In order to introduce contrast on the object surface, we project a random dot pattern using a slice projector from a distance of about 1m. The resulting contrast facilitates the use of depth-from-focus and stereo matching. Figure 5 shows partial shapes of the object from 4 directions.

Initial calibration of the the vision system using Tsai's algorithm gives a translation matrix

$$T_s = (-38.5, 0.0, 833.0)^T .$$

Refining the translation matrix for registration is done for the 'Head' object for all view from V_0 to V_3. Figure 6 shows x and z components of error in the translation matrix T_s with respect to the iteration number. Figure

Figure 5: Partial shapes from 4 views

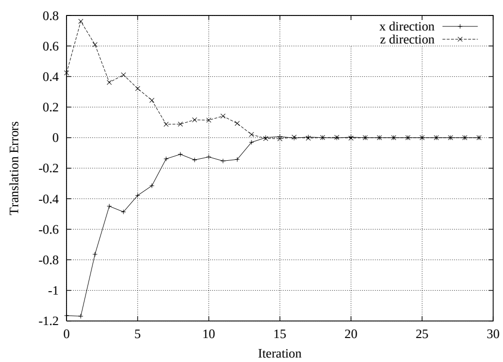
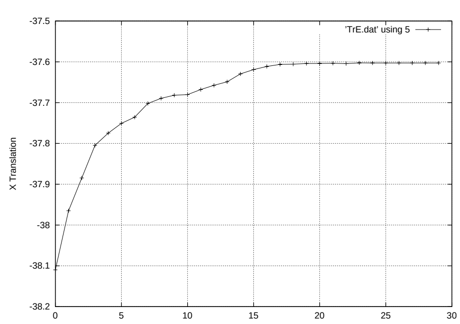
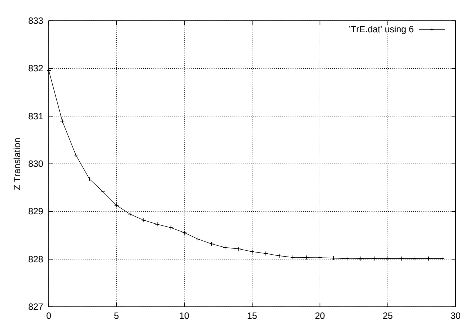


Figure 6: x and z direction's translation error



(a) x direction



(b) z direction

Figure 7: Convergence of T_s

7 shows the actual values of the translation matrix converging with iterations.

After the error minimization, the refined translation matrix is given by

$$T_s{}' = (-37.6, 0.0, 828.0)^T \,.$$

Integration of the 4 partial shapes after the registration give a result as shown in Figure 8. Reconstructed model has total 26717 vertices and 51491 triangles before the mesh optimization.

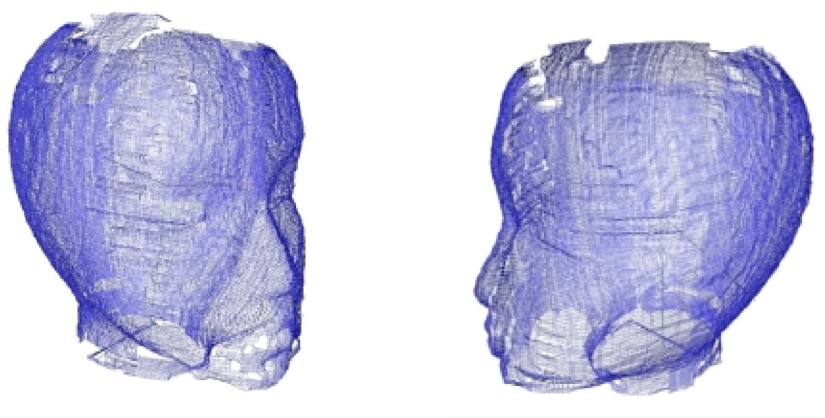


Figure 8: Reconstructed surface model 'Head'

6. CONCLUSIONS

This paper describes a new method for complete 3D shape reconstruction from multiple partial shapes acquired from a digital stereo camera. We take stereo image pairs with parallel stereo geometry. A multi-resolution stereo matching method based on Gaussian pyramid is used. We take stereo pictures of an object from 4 directions of view by rotating the object placed on a rotating stage. Four partial 3D shapes of the object are computed and initially registered using calibration parameters determined by Tsai's method. The translation parameters are then refined by iteratively minimizing the registration error.

We have introduced new algorithms for registration and integration of partial 3D models. The algorithms exploit the epipolar geometry between different views. Correspondence between 3D points are established by projecting overlapping contour segments from one view to another view and finding the closest point on the epipolar line. The average translation error in registration is minimized iteratively.

Integration is done slice by slice using linked lists representing points and segments of contours of object cross-sections. The linked lists for different views are merged based on two closeness criteria to obtain closed contours representing complete cross-sections of the object. Meshes are created from the closed contours to generate a complete model of the object. Results for a real object are presented.

Future research will investigate the use of image texture data in addition to shape data for registering and integrating partial shapes from different views. Image texture data can be projcted from the image plane of one view to the image plane of another view, and stereo image matching techniques can be used to establish correspondence.

Acknowledgments

The support of this research in part by Olympus Optical Corporation is gratefully acknowledged.

REFERENCES

1. I. Söderkvist, "Introductory Overview of Surface Reconstruction Methods," *Lulea University of Technology, Dept. of Mathematics, Reseach Report 10*, 1999.
2. P. Besl and H. McKay, "A Method for Registration of 3D Shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, pp. 239-256, 1992.
3. P. Pritchett and A. Zisserman, "Matching and Reconstruction from Widely Separated Views," *European Workshop SMILE'98*, pp.78-92, 1998.
4. Y. Chen and G. Medioni, "Object Modeling by Registration of Multiple Range Images," *Image and Vision Computing*, Vol. 10, No. 3, pp. 145-155, 1992.
5. H. Gagnon, M. Soucy, R. Bergevin, and D. Laurendeau, "Registration of Multiple Range Views for Automatic 3D Model Building," In*Proceedings of IEEE Computer Vision and Pattern Recognition Conference*, pp. 581-586, June 1994.
6. S. Weik, "Registration of 3D Partial Surface Models Using Luminance and Depth Information," *International Conferences on Recent Advances in 3D Digital Imaging and Modeling*, pp. 93-100, 1997.
7. R.Y. Tsai, "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Camera and Lenses," *IEEE Journal of Robotics and Automation*, Vol. 3, No. 4, Aug. 1987.
8. P.J. Burt, "The Laplacian Pyramid as a Compact Image Code," *IEEE Trans. on Communications*, Vol. 31, No. 4, Apr. 1983.
9. E. Trucco ad A. Verri, *Introductory Techniques for 3D Computer Vision*, Prentice Hall, New Jersey, 1998.
10. B. Curless and M. Levoy, "A Volumetric Method for Building Complex Models from Range Images," In *Proceedings of SIGGRAPH*, pp. 303-312, 1996.
11. A. W. Fitzgibbon, G. Cross, and A. Zisserman, "Automatic 3D Model Construction for Turn-Table Sequences," *European Workshop SMILE'98*, pp.155-170, 1998.