A Vision System for Fast 3D Model Reconstruction

Huei-Yung Lin and Murali Subbarao

Department of Electrical and Computer Engineering
State University of New York at Stony Brook
Stony Brook, New York 11794-2350
{hylin, murali}@ece.sunysb.edu http://www.ece.sunysb.edu/˜cvl

Abstract

A desktop vision system is presented for complete 3D model acquisition. It is fast, low-cost, and accurate. Partial 3D shapes and texture information are acquired from multiple viewing directions using rotational stereo and shape from focus (SFF). The resulting range images are registered to a common coordinate system and a surface representation is created for each range image. The resulting surfaces are integrated using a new algorithm named Region-of-Construction. Unlike previous approaches, the Region-of-Construction algorithm directly exploits the structure of the raw range images. The algorithm determines regions in range images corresponding to non-redundant surfaces which can be stitched along the boundaries to construct the complete 3D surface model. The algorithm is computationally efficient and less sensitive to registration error. It also has the ability to construct complete 3D models of complex objects with holes. A photo realistic 3D model is obtained by mapping texture information onto the complete surface model representing 3D shape. Experimental results for several real objects are presented.

1. Introduction

Three-dimensional (3D) model reconstruction of an object with geometric and photometric information is useful in many practical applications. Often the 3D models are created manually by users. This process is usually time-consuming and expensive. Therefore, techniques to obtain object model automatically from real objects could have great significance.

The reconstruction of a 3D model consists of four key stages [3]: (i) data acquisition, (ii) registration, (iii) surface integration, and (iv) texture mapping. In this paper, we present a complete system that includes all these four stages to create a photo realistic 3D model. The object is placed on a rotation stage in front of a stationary camera. Input image sequences from different viewpoints are acquired by rotating the stage with known rotation angles. The rotation matrix and translation vector of the rotation axis is calibrated using a simple 3-point algorithm and refined during surface integration. In the data acquisition stage, partial 3D shapes and the corresponding focused images are recovered using *rotational stereo* and *shape from focus*. The range images are then registered to the camera coordinate system according to the rotation axis and their viewing directions.

A new surface integration algorithm named *Region-of-Construction* is used to *stitch* the partial 3D models from different viewpoints. Unlike previous approaches [6, 8], the raw data of range images is directly accessed to create non-overlapping regions for integration. It takes advantage of the known topology of each Region-of-Construction to perform fast triangulations. Because continuous regions from same range images are used for integration, the registration error is always limited to the boundaries of Region-of-Construction. This algorithm is also extended to construct a complex object with holes. Holes are reconstructed from the range images using data from two *opposite* viewpoints facing the hole. The integration method is computationally efficient in the sense that no searching is required for mesh triangulation. Finally, the focused images recovered by SFF are mapped onto the reconstructed wireframe model to create a photo realistic 3D model.

2. Data Acquisition

The acquisition of range data and the corresponding texture information is performed by our SVIS-2 camera system [7]. It includes a digital camera, a motorized rotation stage and a PC. The object is placed on the rotation stage and images from different viewpoints are taken by rotating the object with known angles. For each viewpoint, range and focused images are obtained using *rotational stereo* and *shape from focus* [10]. Two sequences of images with different focus positions are taken with a small rotation angle to obtain stereo image pairs. Each sequence of images is used to construct the focused image and a rough depth map using SFF. A more accurate 3D shape is then obtained using rotational stereo on the focused image pair and rough depth maps.

2.1. Rotational Stereo Model

The rotational stereo model used in data acquisition is shown in Figure 1. The rotation axis is described by the unit vector $\vec{n} = (n_1, n_2, n_3)^T$ and the translation vector $\vec{d} = (d_1, d_2, d_3)^T$ in the camera coordinate system. The image pair used for stereo matching is obtained by rotating the object an angle θ with respect to the rotation axis. Let (x_1, y_1, z_1), (x_2, y_2, z_2) be the same object point before and after rotation, respectively; and $(\hat{x}_1, \hat{y}_1)$, $(\hat{x}_2, \hat{y}_2)$ denote the corresponding image point. The images taken before and after rotation are referred to as the first and second image. For each point $(\hat{x}_1, \hat{y}_1)$ in the first image, the corresponding *epipolar line* in the second image is calculated and used for stereo matching. Let $\vec{n}$ be a unit vector along the selected rotation axis and θ be the specified rotation angle about this axis. As shown in [11], the rotation matrix can be written as

$$\mathbf{M}_R(\theta) = (1-\cos\theta)\begin{bmatrix} n_1^2 & n_1 n_2 & n_1 n_3 \\ n_2 n_1 & n_2^2 & n_2 n_3 \\ n_3 n_1 & n_3 n_2 & n_3^2 \end{bmatrix} + \sin\theta \begin{bmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{bmatrix} + I\cos\theta$$

The rotation matrix for any rotation axis with a translation vector $\vec{d}$ is then

$$\mathbf{R}(\theta) = \mathbf{T}^{-1} \cdot \begin{bmatrix} \mathbf{M}_R(\theta) & 0 \\ 0 & 1 \end{bmatrix} \cdot \mathbf{T} \tag{1}$$

where

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & d_1 \\ 0 & 1 & 0 & d_2 \\ 0 & 0 & 1 & d_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

By perspective projection, the relationship between an object point (x_i, y_i, z_i) and an image point $(\hat{x}_i, \hat{y}_i)$ is given by

$$x_i = \hat{x}_i t_i, \qquad y_i = \hat{y}_i t_i, \qquad \text{and} \qquad z_i = f t_i$$

where t_i's are unknown parameters. Thus, we have the equations

$$\begin{pmatrix} \hat{x}_2 t_2 \\ \hat{y}_2 t_2 \\ f t_2 \\ 1 \end{pmatrix} = \mathbf{R}(\theta) \cdot \begin{pmatrix} \hat{x}_1 t_1 \\ \hat{y}_1 t_1 \\ f t_1 \\ 1 \end{pmatrix} \tag{2}$$

for the same object point projected on different images. Solving equations (1) and (2), the epipolar line on the second image for any fixed $(\hat{x}_1, \hat{y}_1)$ is given by

$$\hat{y}_2 = m \cdot \hat{x}_2 + c \tag{3}$$

where m and c are functions of $\hat{x}_1$ and $\hat{y}_1$.

In our rotational stereo model, the stereo matching is done along the epipolar line derived above at 16×16 image block intervals for only the foreground pixels as determined by SFF. Therefore image rectification is avoided.

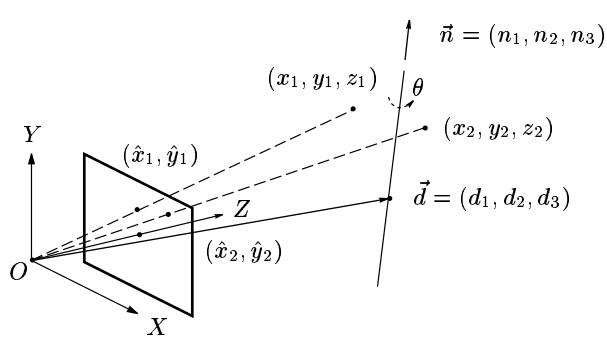

Figure 1: Rotational stereo model

Figure 2: SVIS-2 camera system

2.2. Implementation

The rotational stereo with SFF is implemented on our Stonybrook VIsion System 2 (SVIS-2). The object is mounted on the rotation stage and the camera is placed in front of it such that the optical axis is close to the rotation axis. A simple 3-point calibration method is used to estimate the rotation axis. Three images of a planar object are taken with different rotation angles, 0, θ and 2θ degrees. Three fixed points with known distances between each other are used to determine their 3D points. One of the 3D points and its corresponding points after θ and 2θ rotation can be used to find the plane perpendicular to the rotation axis. Together with the fact that the rotation axis passes through the center of the circumscribed circle of the triangle, the rotation axis can be uniquely determined. Since the rotation

matrix is unknown, the corresponding image points are used to identify their object points:

$$\|z_1\mathbf{p}_1 - z_2\mathbf{p}_2\| = \|z_2\mathbf{p}_2 - z_3\mathbf{p}_3\| = 2\cos\frac{\theta}{2}\|z_1\mathbf{p}_1 - z_3\mathbf{p}_3\|$$

where $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$ are the image points in different images associated with the same object point after rotations of 0, θ, 2θ degrees, and z_1, z_2, z_3 are the corresponding depths to be solved.

In the experiments, we use objects with fine texture to help stereo matching. Two sequences of 4 images are recorded with different focus settings before and after rotating the object by 6 degrees for each viewpoint. Shape from focus is applied on each sequence of images to get the focused image and a rough depth map. 16×16 image blocks are used to obtain a 80×60 rough depth map and a 1280×960 focused image. The depth map is thresholded to segment the focused image into foreground and background regions.

Rotational stereo analysis is then carried out using the focused images and the initial depth map estimated by SFF to get an accurate depth map. Sum-of-squared-difference measure on 16×16 image blocks is used for matching in the foreground regions. A fast stereo matching is done by limiting the correspondence search to a small segment on the epipolar line determined by the rough depth map obtained from SFF. The length of the segment for searching is computed using the maximum expected error of SFF. Finally, the 3D shape is obtained by an inverse perspective projection of the resulting depth map. This data acquisition procedure is repeated for 4 viewpoints by rotating the object every 90 degrees. The output of 3D shapes with different resolution settings are shown in Figure 3.

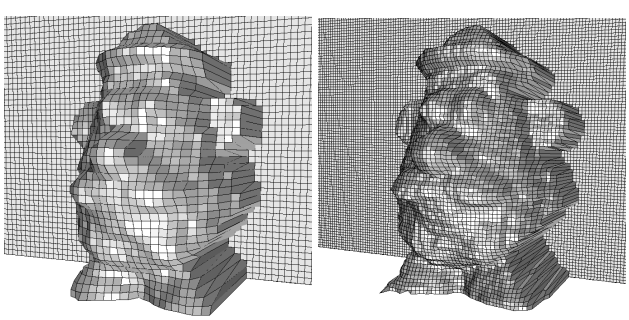

Figure 3: Different resolutions of a 3D shape (left: 80×60, right: 160×120).

3. Registration

To create a complete description of an object, multiple range images are registered to a common coordinate frame. The background points of each range image are eliminated before registration and only the object (foreground) points are registered to a common coordinate system.

Our registration method consists of two steps. First, we find the initial estimated transformation by calibrating the rotation axis described in the previous section. The translation and unit vector of the rotation axis are used to compute the rotation matrix $\mathbf{R}(\theta)$ between the coordinate system of the viewpoint at θ degrees and the camera coordinate system by Eq. (1). The range images are registered to the camera coordinate system using the inverse transformation $\mathbf{R}(\theta)^{-1}$.

Second, the estimated registration is further improved during surface integration. The overlapping parts of range images are used to refine the rotation axis and increase the accuracy of registration. We search a small neighborhood of the calibrated rotation axis to find the one with minimized error on the overlap of consecutive range images in the least squared sense. This refined rotation axis is then used to compute the new rotation matrix and translation vector for registration. This refinement step provides good results with much less computation than other registration techniques such as Iterative Closest Points algorithm (ICP) [2, 4].

4 Surface Integration

Given a set of registered views, an integration algorithm should combine the partly overlapping data sets into a complete non-redundant 3D data set without any loss of detail in the original raw data. In the existing literature, Hoppe *et al.* [6] and Amenta *et al.* [1] construct 3D surface from unorganized points. Hilton *et al.* [5], Soucy and Laurendeau [9] use structured data to combine multiple range images. All of them assume that accurate range data is available and they are perfectly registered. It is difficult to obtain such data in a practical data acquisition system. Therefore, an integration algorithm which is robust under the presence of noise and registration error is an important goal in this paper.

4.1. Region-of-Construction Algorithm

Our input depth-map data is given on a regular grid of points for each partial 3D model. Therefore we develop a specialized triangulation method depending on the viewpoints. The basic idea is to *stitch* the *Regions-of-Construction* of different viewpoints at their boundaries to create a complete 3D model. This algorithm takes advantage of the known topology of each range image and does not involve any spatial search of the data points. The mesh triangulation is done using indices of data points on a grid network. As a result, it is much faster than the general algorithms mentioned earlier.

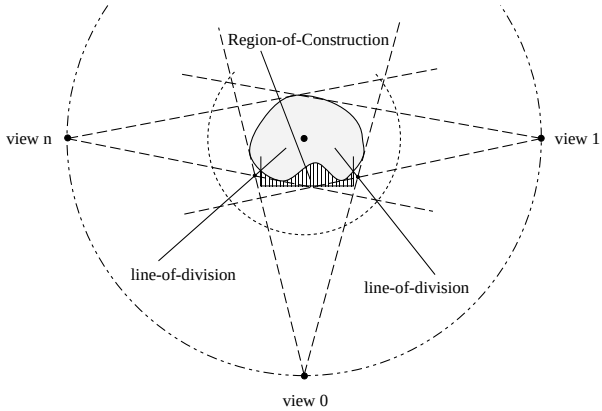

Figure 4: Region-of-Construction on a cross section.

The overlapping part of range images from two consecutive viewpoints is divided into two regions and each region is assigned to one of the range images for construction. We first create the *Region-of-Construction* for each range image and then use them to create a non-overlapping data set for surface integration. Region-of-Construction is defined in Figure 4. In the figure, the range image corresponding to the shaded region is the Region-of-Construction for view 0. The left (right) *line-of-sight* is defined as the line determined by the viewpoint and the leftmost (rightmost) data point. The angular bisector of right (left) line-of-sight of $i-1$ ($i+1$) range image and left (right) line-of-sight of i range image is defined as the left (right) *line-of-division* for range image i. The 2D Region-of-Construction for each cross section is then bounded by the left and right line-of-division. For each viewpoint, the Region-of-Construction includes all the 2D cross sections. For each Region-of-Construction, the triangular mesh is created as follows. We start at the upper left corner of the range image, find the points belonging to the Region-of-Construction, mark them as *valid points* and establish the connection. For each valid point there exist five possible tessellations for triangles or quadrilaterals (see Figure 5). For the first and second cases, the column index of the first valid point in the ith row is smaller or larger than the column index of the first valid point in the jth row, where $|i-j|=1$. Without loss of generality, assume row i contains the smaller column index. Let p be the smaller and q be the larger column index, then we make triangles by connecting indices $(j,q)-(i,p)-(i,p+1)$, $(j,q)-(i,p+1)-(i,p+2)$, ..., until $p=q-1$. For the third and fourth cases, the column index of the last valid point in the ith row is smaller or larger than the column index of the last valid point in the jth row, where $|i-j|=1$. Without loss of generality, assume row i contains the smaller column index. Let p be the smaller and q be the larger column index, then one or more triangles are created by connecting indices $(i,p)-(j,p)-(j,p+1)$, $(i,p)-(j,p+1)-(j,p+2)$, ..., $(i,p)-(j,q-1)-(j,q)$. In these four cases one or more triangles are created depending on the column index difference between any two consecutive rows. For the last case, the column index of the valid point in the ith row is the same as the column index of the valid point in the $(i+1)$th row. In this case, lots of quadrilaterals in the central part of the range image are produced and the number of quadrilaterals depends on the difference between the last and first column index. Those quadrilaterals are further divided into triangles by connecting the shorter diagonals.

After the meshes for Regions-of-Construction are created for each view, the partial 3D models are *stitched* together by connecting the last valid point of the current range image to the first valid point of the next range image in the same row (see Figure 5). The resulting quadrilaterals on the boundaries of two range images are also broken into triangles with shorter diagonals.

4.2. Complex Object with Holes

Since the Region-of-Construction algorithm provides only non-overlapping surfaces, complex objects with holes can be reconstructed by selecting proper viewpoints that include the holes. We assume that if a hole can be observed from one viewpoint, it can also be observed from the *opposite* viewpoint. Generally this holds for most real objects. Under this assumption, a 3D model is first constructed using the previous algorithm without considering its hole. Then the boundary points of the hole which belongs to two opposite viewpoints are connected using the following algorithm.

First, we connect the top and bottom rows of two range images respectively. The row index can be different. The column indices of the top (bottom) rows are used to create a triangle/quadrilateral mesh similar to the previous

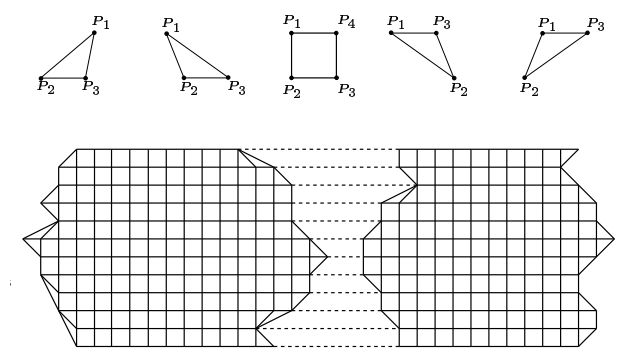

Figure 5: Five possible tessellations within a Region-of-Construction (top) and *stitching* between them (bottom).

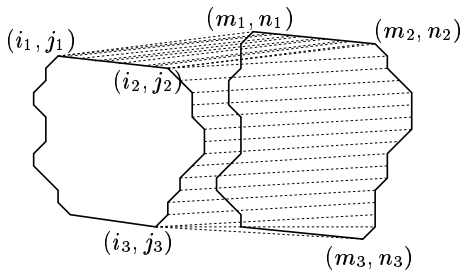

Figure 6: Mesh for creating a hole. The index differences between (m, n) and (i, j) are used to generate quadrilaterals or triangles.

section. For a side boundary of the hole, another triangle/quadrilateral mesh is created with variable row index and fixed column index (one belongs to object points and next to a background point). Figure 6 illustrates this *hole-creating* process. The quadrilaterals are further broken into triangles with short diagonals.

In some cases the observed hole does not appear in any Region-of-Construction defined in the previous section but it gives depth discontinuities in one of the range images (see Figure 9). We define the *principal view* as the viewpoint containing the hole. The lines-of-division of principal views are shifted such that the new Regions-of-Construction can completely cover the hole.

5. Texture Mapping

One major advantage of using images (instead of say laser ranging) to acquire 3D model is that the recorded images are used not only for measuring the 3D shape of objects but also for providing the texture information. The focused images used for texture mapping are obtained from SFF with different focus positions. For each Region-of-Construction the texture map is extracted from the focused image of its corresponding viewpoint. The texture map at the boundary strip connecting two range images is obtained by projecting the vertices at the boundary of one range image to the other intensity image and extracting the texture from that intensity image.

6. Experimental Results

The above algorithms were tested on a number of real objects. Figure 7 shows the wireframe and texture mapped model of a toy object. Figure 8 shows the result of a head object which contains 26,977 vertices and 53,687 polygons. In Figure 9, the object contains a hole close to boundaries of the range image. The Region-of-Construction of principal views are increased (illustrated in wireframe) in order to completely cover the hole.

It takes approximately 15 minutes to acquire stereo image pair with 4 different focus positions for 4 views. The execution time for creating complete 3D models– shape from focus (SFF), stereo matching (SM), surface integration (SI) and texture mapping (TM) are shown in Table 1 for several objects (all in seconds, on a Pentium II 450 MHz PC).

Table 1: Execution times several objects.

object	SFF	SM	SI	TM	Total
toy	5.5	74.2	1.3	4.9	85.9
head	5.7	115.2	1.9	6.7	129.5
detergent	3.0	37.7	0.5	5.4	46.6

Our 3D model reconstruction uses the raw data from range images without any modification. The accuracy depends on the accuracy of range images (rotational stereo and SFF) and registration (rotation axis calibration). To measure the overall accuracy of our system, we use a cylinder as test object. The acquired 3D data set is fitted to the cylinder from the physical measurement to calculate the average error. In our experiment, the average error for each data point was found to be 0.28% (0.44 mm error for a cylinder of diameter 155 mm) at a distance of about 830 mm from the camera.

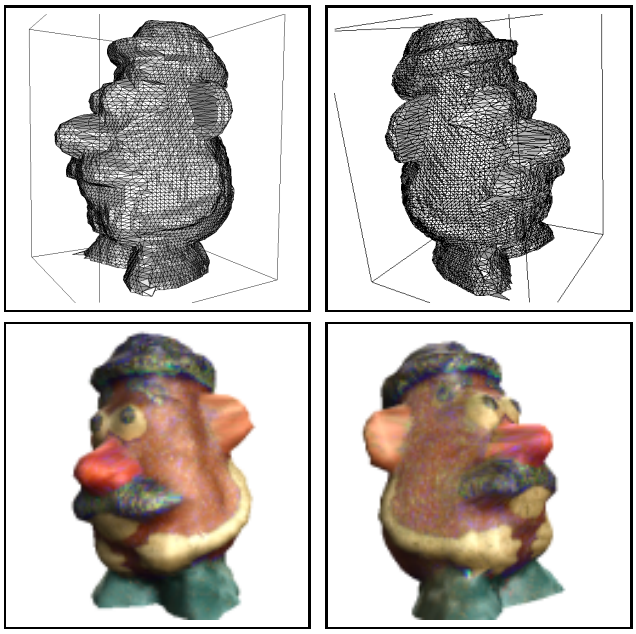

Figure 7: Wireframe and textured models of a toy object.

7. Conclusion

We have designed and implemented a digital vision system for automatic 3D model reconstruction. The system

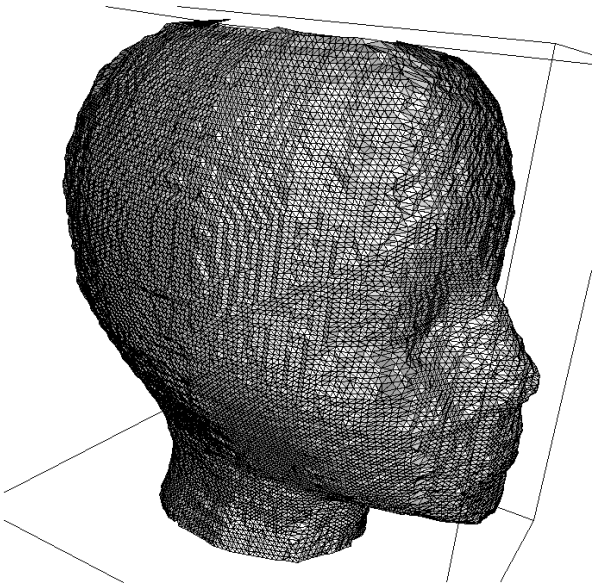

Figure 8: Wireframe model of the head object.

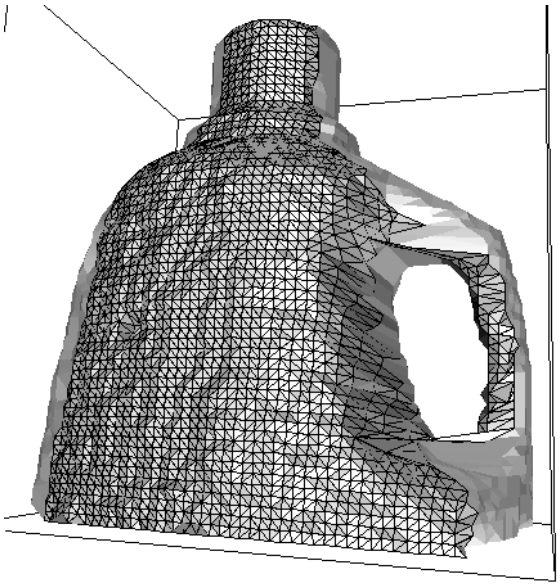

Figure 9: Complex object with a hole, one of the Region-of-Construction is shown in wireframe.

is comprehensive in that it includes all stages– data acquisition, registration, surface integration, and texture mapping, to create a photo realistic 3D model. The complete 3D model is constructed by merging multiple range images and mapping the texture information acquired by rotational stereo and SFF. A new surface integration algorithm based on Region-of-Construction is developed for fast 3D model reconstruction. It is also capable of constructing complex objects with holes. Experimental results have been presented for several real objects. Our system can be implemented with low-cost equipment and constructs complete 3D models in under 30 minutes. Future research will focus on extending the Region-of-Construction algorithm from the viewpoint-based regions to the regions with best viewing directions.

Acknowledgments

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

References

[1] N. Amenta, M. Bern, and M. Kamvysselis, "A new Voronoi-based surface reconstruction algorithm," *SIGGRAPH 98*, pp. 415-421, 1998

[2] P. J. Besl and N. McKay, "A method of registration of 3D shape," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 2, pp.239-256, Feb. 1992.

[3] Y. Chen and G. Medioni, "Object modeling by registration of multiple range images," *Image and Vision Computing*, Vol. 10, No. 3, pp. 145-155, March 1992.

[4] C. Dorai, J. Weng and A. K. Jain, "Optimal registration of object views using range data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 10, pp. 1131-1138, Oct. 1997

[5] A. Hilton, J. Stoddart, J. Illingworth, and T. Windeatt, "Reliable surface reconstruction from multiple range images," *Proceedings of European Conference on Computer Vision '96*, pp. 117-126, 1996

[6] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle, "Surface reconstruction from unorganized points," *ACM Computer Graphics*, Vol 26, pp. 71-78, July 1992.

[7] H. Y. Lin and M. Subbarao, "Three-dimensional model acquisition using rotational stereo and image focus analysis," *Proceedings of SPIE*, Vol. 4189, pp. 201-210, Nov. 2000.

[8] M. Reed and P. K. Allen, "3D modeling from range imagery," *Image and Vision Computing*, 17(1):99-111, Feb. 1999.

[9] M. Soucy and D. Laurendeau, "A general surface approach to the integration of a set of range views," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 17, pp. 344-358, April 1995.

[10] M. Subbarao and T. S. Choi, "Accurate recovery of three-dimensional shape from image focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, pp. 266-274, March 1995.

[11] E. Trucco and A. Verri, *Introductory Techniques for 3D Computer Vision*, Prentice Hall, New Jersey, 1998.