A Digital Vision System for Three-Dimensional Model Acquisition

Ta Yuan, Huei-Yung Lin, Xiangdong Qin, Murali Subbarao
Computer Vision Laboratory
Department of Electrical and Computer Engineering
State University of New York at Stony Brook
Stony Brook, New York 11794-2350, USA

ABSTRACT

A digital vision system and the computational algorithms used by the system for three-dimensional (3D) model acquisition are described. The system is named Stonybrook VIsion System (SVIS). The system can acquire the 3D model (which includes the 3D shape and the corresponding image texture) of a simple object within a 300 mm × 300 mm × 300 mm volume placed about 600 mm from the system. SVIS integrates Image Focus Analysis (IFA) and Stereo Image Analysis (SIA) techniques for 3D shape and image texture recovery. First, 4 to 8 partial 3D models of the object are obtained from 4 to 8 views of the object. The partial models are then integrated to obtain a complete model of the object. The complete model is displayed using a 3D graphics rendering software (Apple's QuickDraw). Experimental results on several objects are presented.

Keywords: 3D object modeling, multiple view registration, multiple view integration, image focus analysis, stereo image analysis, range image.

1. INTRODUCTION

The 3D model of an object consists of two types of information – (i) the 3D shape of the object, and (ii) the image texture on the outer visible surface of the object. Recovering the first type of information (3D shape) is a difficult problem in the computer vision area.[2,1] Some popular techniques are stereo, shading, focus analysis, structured light analysis, etc. As for the second type of information (image texture), it is recovered easily from the image recorded by a camera.

Stereo Image Analysis (SIA) is perhaps the most widely used technique for 3D shape recovery in computer vision.[2,1] However, this technique has some inherent computational problems (e.g. correspondence and occlusion) related to matching stereo image pairs. In our earlier work,[7,8] we mitigated these problems by integrating Image Focus Analysis (IFA)[3–5] and Image Defocus Analysis (IDA)[6,5] with SIA. In the current work, IDA which can estimate the approximate range of an object was not used because the range was known (600 mm to 900 mm). In the current work, IFA first provides an approximate 3D shape of the object which simplifies the stereo matching problem in SIA. The approximate 3D shape is refined by SIA to obtain a more accurate 3D shape.

In this paper we extend our previous work[7,8] on recovering the 3D shape and image texture of a single view of an object in two main ways– computational algorithms and hardware architecture. The stereo image matching algorithms have been improved to obtain more accurate partial 3D models of objects. New computational algorithms are provided for representing and integrating partial 3D models obtained from different views into one complete 3D model of the object. Integrating partial models involves several steps. First, the measured coordinates of points on the surface of the object provided by partial models are transformed to an object centered cylindrical coordinate system. This requires calibrating the rotation axis of the object. A procedure for this purpose has been provided. The transformed coordinates of different partial models are merged by taking rotation angles of the partial models into account. This results in a set of points corresponding to discrete sampling of the complete object's surface. These discrete sample points are used to interpolate the surface. Then the surface is resampled at regular intervals to model the surface with quadrilateral surface patches. The vertices of the quadrilaterals are projected back onto

Further author information:
Email: {tyuan,hylin,xdqin,murali}@ece.sunysb.edu
WWW: http://www.ece.sunysb.edu/~cvl

the focused images along different directions of view to obtain the image textures of the corresponding quadrilateral surface patches. The complete 3D model is then printed to a file in the 3D metafile format (3dmf) suitable for rendering by the Apple's QuickDraw 3D Viewer software.

The earlier computing and image digitization hardware of SVIS[7,8] have been completely replaced with new and more powerful ones (Fig. 1). A computer controlled motor system has been added to rotate the object by known amounts so that different views of the object are automatically obtained.

At present, SVIS works well for simple objects defined as objects whose cross-sections perpendicular to their rotation axis can be defined in polar coordinates (with its origin at the rotation axis) by a function of the form $r(\theta)$. SVIS will be extended to work with more complicated objects in the future. The results of 3D model acquisition for several objects are presented.

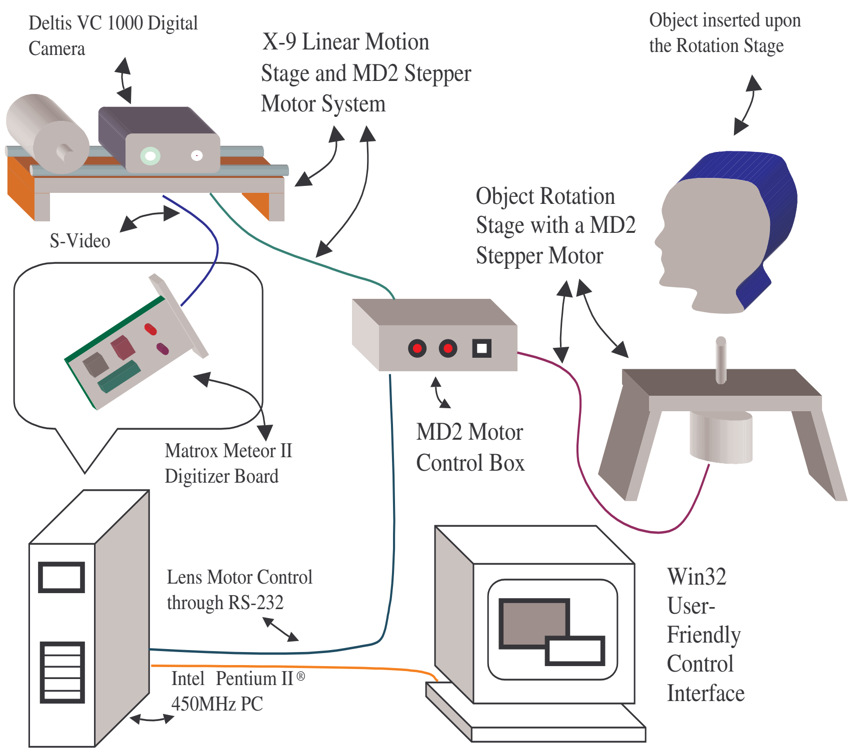


Figure 1. Stonybrook VIsion System (SVIS)

2. ACQUISITION OF 3D SHAPE AND FOCUSED IMAGE

The method used here for measuring the 3D shape and focused image of a single view of an object is basically similar to our earlier work described in.[7,8] Here we leave out the details of our earlier method, but include all the important extensions to it.

In the experiments, the parameters of SVIS were adjusted for objects that fit inside a 300 mm $\times$ 300 mm $\times$ 300 mm cube placed at a distance of about 600 mm from the camera system (Fig. 2). The objects are assumed to be simple in the sense explained earlier. Images were recorded by a digital camera (Olympus DELTIS VC 1000). The camera focal length was set to 19.6 mm, and F-number to 4. Different focus settings were obtained by moving the camera's motorized lens by a Personal Computer (PC).

In the first step, SVIS uses 5 images recorded with different focus settings in IFA. The distance range of 600 mm to 900 mm for object used in the experiments corresponds to focus settings of 113 to 105 steps for the camera's lens position. Therefore the images were acquired at 2 lens step intervals. At this stage, although color images are recorded, for the sake of computational efficiency, gray-level images are used in IFA. The gray level images were computed from color data as the mean value of R, G and B components. This step results in an initial estimate for 3D shape and focused image.

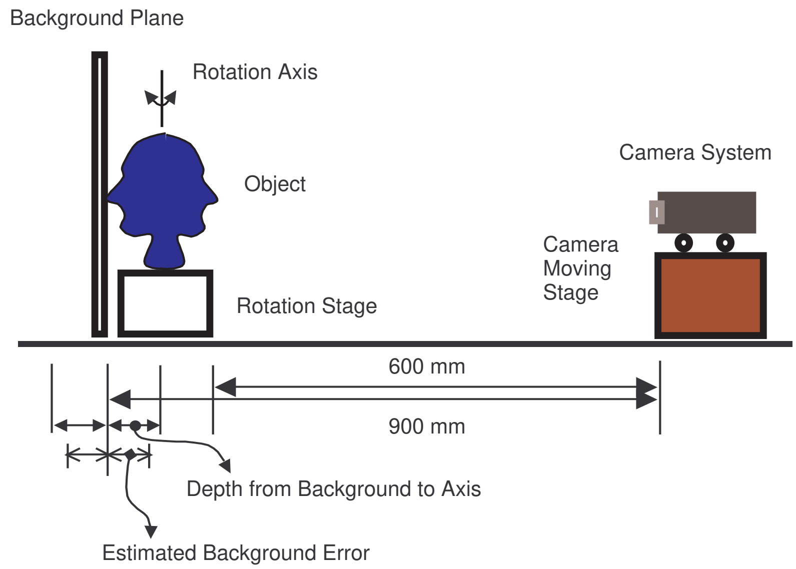


Figure 2. Experimental setup

For each view of the object, a 432×432 subimage was extracted from a 640×480 image recorded by the camera. The center of the 640×480 recorded was taken to be the point where the optical axis intersected the image plane. It was also the origin of the image coordinate system for perspective projection. IFA was applied in 16×16 image blocks to obtain a 27×27 depth-map and a 432×432 focused image. The depth-map is thresholded to segment both the depth-map and the focused image into two regions, one corresponding to background region (points farther than the expected distance of object points) and object or foreground region. This above procedure is repeated for both the left and right stereo camera positions.

The initial depth-map estimate obtained by IFA is refined using SIA with color image data. The stereo baseline is 50 mm. The single digital camera is moved by a motor perpendicular to its optical axis to create the effect of a stereo camera. Focused images recovered by IFA corresponding to left and right stereo images are used in matching. Matching is done for 16×16 image blocks using the sum-of-squared-difference measure for color images.[8] In the matching step, only the foreground regions of the left and right stereo images are considered. This saves a lot of computation.

Matching is done by considering 16×16 image blocks (in the foreground region) from the right focused image and searching for the matching image (foreground) regions in the left focused image. The epipolar line obtained by calibration was almost horizontal (parallel to the $\hat{x}$-axis on the image plane). The search was limited to a 5 pixel wide band in the vertical direction (i.e. perpendicular to the epipolar line). Along the epipolar line, the search was further limited by using the depth-map obtained by IFA and the maximum expected depth error in the results of IFA (4 lens steps in focus setting).

The matching was further improved by detecting mismatches due to occlusion. This was done as follows. For each 16×16 foreground image block (Block 1 in Fig. 3) in the right focused image, the best matched foreground image region (Block 2 in Fig. 3) in the left focused image was found. This was done by minimizing the SSD for color image data along the epipolar line. Then the best matching image region in the left focused image (i.e. Block 2) was back matched by searching for the best match (Block 3 in Fig. 3) in the right focused image (along the epipolar line). If the distance between the centers of Block 1 and Block 3 in Fig. 3 was more than half the size of the matching block (i.e. $16/2 = 8$ pixels), then Block 1 was considered to be invisible (occluded) from the left camera. The disparity for that block was taken to be the same as the next or previous image block (along the epipolar line). The final result of this step is a 27×27 depth-map array segmented into object and background points, and a corresponding 432×432 focused image.

The visible surface from the camera's view point is modeled in two parts– (i) the 3D shape of the surface specified by the 27×27 depth-map array, and (ii) the image texture of the surface specified by the focused image recovered by IFA. This constitutes a partial model of the object as only a part of the object is visible from a given direction of view. The 3D shape of the surface is modeled as follows. The camera coordinates of the points in the depth-map are obtained by an inverse perspective projection for the camera and printed as (X_i, Y_i, Z_i) triples for $i = 1, 2, ..., n$, where n is the total number of points ($n = 27 \times 27$). These triples, constitute a list of vertices in the 3D space of camera coordinate system, where the vertices are uniquely identified by their indices i. The rectangular grid specified by the 27×27 array is used to create a list of quadrilaterals in the 3D space where each quadrilateral corresponds to one rectangle in the grid. The clockwise ordering of corner points of the rectangle in the grid are used to specify the quadrilateral as a list of four ordered vertex indices. For each quadrilateral, the image texture is obtained from the focused image in the corresponding rectangle in the rectangular grid. Thus, for a given direction of view, the partial 3D model of the object is obtained. This model can be displayed on a computer monitor using a rendering software such as Apple's QuickDraw or GeomView of University of Minnesota.

The above method for obtaining partial 3D model of an object is repeated for 4 to 8 different views of the object. Different views of the object are obtained by rotating the object using a computer controlled motor by known angles (45 to 90 degrees).

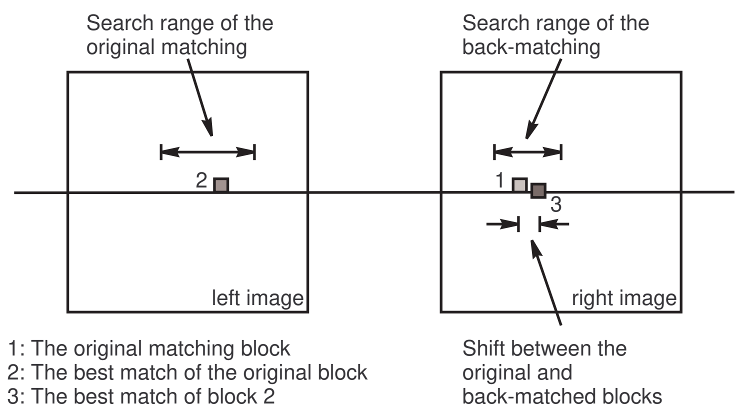


Figure 3. Occlusion detection

The position of the rotation axis is calibrated as follows. First, the experimental set up is arranged initially so that the rotation axis is on or very close to the optical axis of the right camera. Therefore X_0 (Fig. 4) was almost zero (after calibration it was found to be about 25 mm). Also, the distance of the rotation axis from the optical center of the camera was measured physically with a tape. In the experiments, it was approximately 750 mm. These initial estimates were improved through calibration as follows. The motor axle for rotating the object was extended (after dismounting the object) by attaching a thin wire to it. A flat high contrast paper was attached to the axis, and one estimate of the position of wire was obtained using IFA. A better estimate was obtained using SIA. For the bare wire a right stereo image of the wire was recorded after focusing the wire. The wire was almost vertical and parallel to the $\hat{y}$-axis in the image. The $\hat{x}$ coordinates of the wire were recorded at two points, one near the top and another near the bottom of the image. The $\hat{x}$ coordinates were recorded eight times at intervals of 45 degree rotation of the axis. The mean of the eight $\hat{x}$ coordinates was taken to be the $\hat{x}$ coordinate of the actual rotation axis in the right image. This process was then repeated for the left camera position to determine the $\hat{x}$ coordinate of the rotation axis in the left stereo image. Then the Z coordinate of the rotation axis was determined from the disparity of the mean $\hat{x}$ coordinates of the axis in the left and right stereo image pair.

3. INTEGRATING PARTIAL 3D MODELS

In order to simplify integrating the partial 3D models obtained in the previous section, the following assumptions are made.

1. The optical axes of the left and right cameras are perpendicular to the baseline, and parallel to each other. The camera coordinate system is a left handed system at the optical center of the right camera with the X-axis aligned with the baseline and the Z-axis aligned with the optical axis (Fig. 5).
2. The rotation axis is perpendicular to the plane of the baseline and the optical axis.
3. For every cross-section of the object perpendicular to the rotation axis, the rotation axis is inside the cross-section.
4. The object is *simple* in the sense described earlier, i.e. for each cross section of the object perpendicular to the rotation axis, there is a 1-1 correspondence between any point on the contour and the angle θ of the direction vector from the rotation center to that point (Fig 4). In other words, in a cylindrical coordinate system (r, θ, y) with it's y-axis coinciding with the axis of rotation, the contour of the object's cross-section at every y can be specified by a function of the form $r(\theta)$.
5. The rotation angles of the object provided by the computer controlled motor are accurate.

Let the rotation axis intersect the XZ-plane at $(X_0, 0, Z_0)$ in the camera coordinate system (Fig. 5). This point is taken as the origin of the object coordinate system for representing the object. It is a cylindrical coordinate system (r, θ, y) with the axis of the cylinder aligned with the rotation axis (parallel to the Y axis) and the angle θ measured with reference to the z-axis.

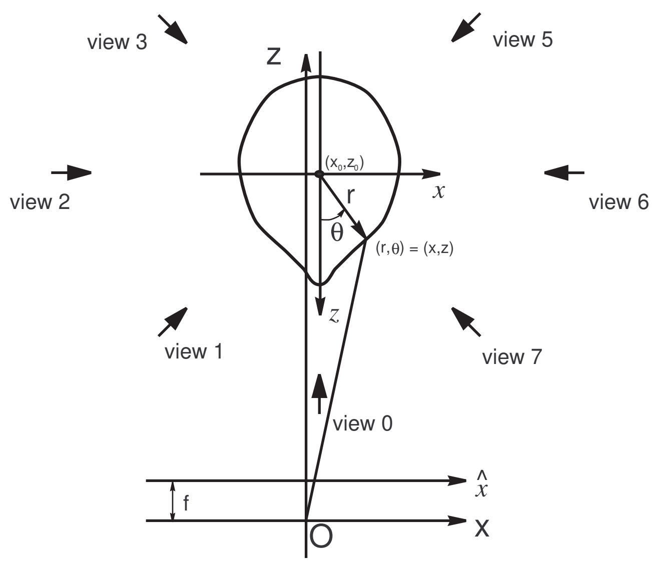


Figure 4. Object's cross-section and viewing directions

In order to integrate partial 3D models of the object to obtain a complete 3D model of the object, several alternative algorithms were considered. After encountering several complications, we arrived at the following algorithm. We will describe the algorithm for the case when eight partial 3D models of the objects were used corresponding to eight views of the object at 45 degree intervals. The views are referred to with their indices $0, 1, 2, ..., 7$.

First, each of the eight partial 3D models represented in the camera coordinates (X, Y, Z) is converted to a representation in the cylindrical object coordinate system (r, θ, y). At this point, only the vertices that are part of the object are retained, and other vertices belonging to the background are eliminated. The background points are eliminated by using a threshold on the Z coordinate of the points. In the experiments, we also eliminated points near the left and right borders of the image (columns 0 to 6 and 21 to 26 in the 27×27 depth-map) as these columns contained points very close to the occlusion boundary of the object. The coordinate (X, Y, Z) of each point is converted to the cylindrical coordinates using the following relations:

$$\begin{aligned} \theta &= \tan^{-1} \frac{X - X_0}{Z_0 - Z} \\ r &= \sqrt{(X - X_0)^2 + (Z - Z_0)^2} \\ y &= Y \end{aligned} \tag{1}$$

The value of θ computed above was adjusted for rotation of the object for different views. If the views are denoted by $n = 0, 1, 2, ..., 7$, and the rotation interval is 45 degrees, then the adjustment for θ in degrees is given by

$$\theta = \theta - n * 45. \tag{2}$$

Having retained only points on the object, then computed cylindrical coordinates of the points, and adjusted their θ for rotation, all the points from different views can be merged into one set. In the object coordinate system, this merged set of points represent a discrete sampling of the visible surface of the object. Since the visible surface is assumed to be simple so that it can be represented by a function of the form $r(\theta, y)$ in the cylindrical object coordinate system, the merged set of points can be thought of as a discrete sampling of this function. Given these discrete samples, we obtain a complete 3D model of the object as follows.

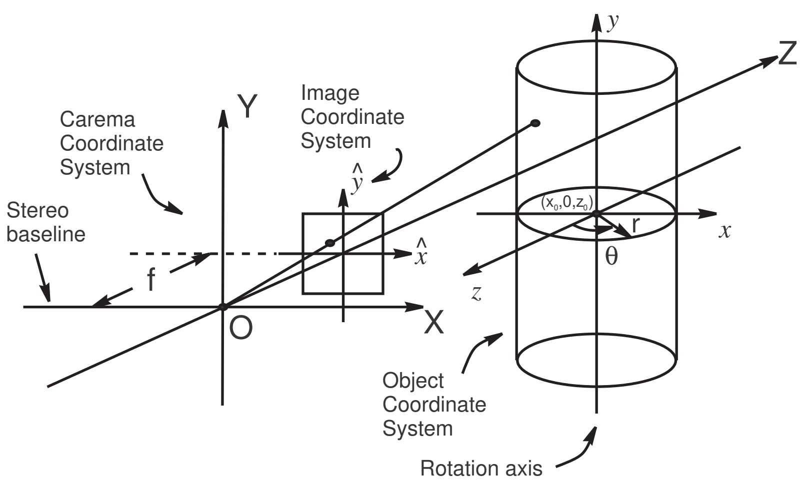


Figure 5. Coordinate systems

The discrete sample points are used to interpolate and uniformly resample the object's surface in the (θ, y) space. In our experiments, the object's surface was uniformly resampled by 27 points in the y space, and 120 points (3 degree intervals) in the θ space. The resulting rectangular sampling grid in the (θ, y) space was used to define a set of vertices (corresponding to sample points) and quadrilaterals (corresponding to rectangles in the grid) that give a piecewise approximation of the object's 3D shape. In order to render this 3D shape on a computer monitor, the coordinates of the vertices were computed in the Cartesian object coordinate system from their cylindrical coordinates as follows:

$$\begin{aligned} x &= r \cdot \sin\theta \\ y &= y \\ z &= r \cdot \cos\theta \end{aligned} \tag{3}$$

In our experiments, we used a simple separable linear interpolation scheme for resampling. First the 27×27 depth-map obtained for each of the eight views of the object was resampled vertically at 27 points along the Y-axis. Then the points on the object were represented in the cylindrical object coordinate system. After this, for each value of y, the surface was resampled at 3 degree intervals (120 points) using a simple linear interpolation scheme.

Having modeled the 3D shape of the object by a set of vertices and quadrilaterals, the image texture of the object is modeled by specifying the image texture of each of the quadrilateral. For each quadrilateral, the corresponding image texture can be computed as follows. The quadrilateral in the object space is projected onto one of the focused images obtained from different directions of view in Fig 4. A vertex at (X, Y, Z) (camera coordinates) of a quadrilateral projects to image coordinates $(\hat{x}, \hat{y})$ in the corresponding focused image given by

$$\begin{aligned} \hat{x} &= \frac{X \cdot f}{Z} \\ \hat{y} &= \frac{Y \cdot f}{Z} \end{aligned} \tag{4}$$

where f is the focal length of the camera. The viewing direction for projecting a quadrilateral is taken to be that which is closest to the mean theta value of the four vertices of the quadrilateral. The image texture of the quadrilateral is the image within the projection area of the quadrilateral on the focused image. The above method of choosing a focused image for projecting the quadrilateral will be bad if the surface normal to the quadrilateral is almost perpendicular to the direction of view. In this case, the quadrilateral will project onto a very small region. Therefore the image texture will be distorted due to coarse sampling. This will show up when the quadirlateral is viewed directly along its surface normal. A better way is to find all direction of views in Fig. 4 from which the quadrilateral is visible, and choose that direction for which the dot product of the surface normal of the quadrilateral and the direction of view is a maximum.

In the experiments, each quadrilateral was first projected onto a focused image by projecting its vertices. The full focused image was a 432×432 image. A rectangular subimage just enclosing the projection of the quadrilateral was extracted (see Fig. 6). The boundaries of the extracted subimage was parallel to the boundaries of the full focused image. The subimage enclosing the projected quadrilateral, and the coordinates of the vertices of the projected quadrilateral on the subimage are used by the 3D metafile format for mapping image texture onto the quadrilateral.

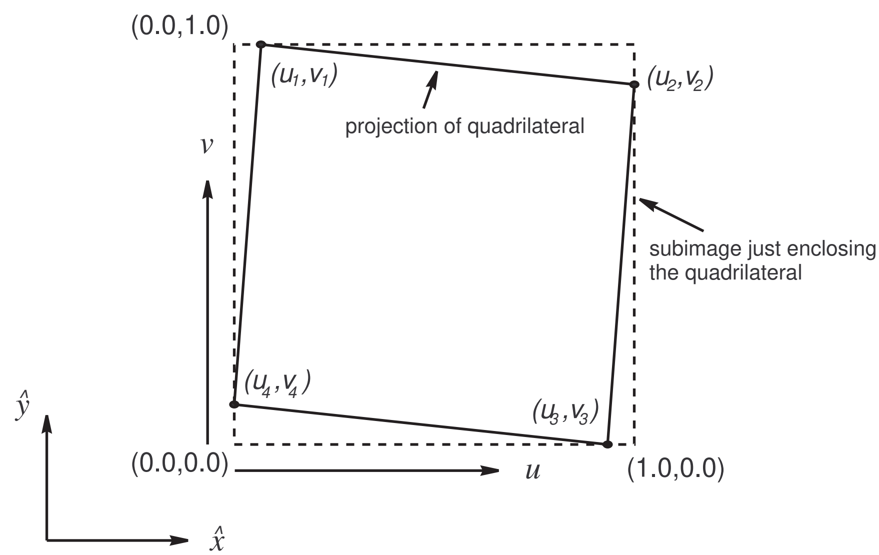


Figure 6. Texture mapping

3.1. Finding the rotation axis

The estimate of the position of the rotation axis obtained in Section 2 can be further improved using the fact that different views of the object have overlapping parts. Suppose that the depth-maps obtained from the i-th view $(i = 0, 1, 2, ..., 7)$ is expressed in the cylindrical object coordinates as $r_i(\theta, y)$. Also, suppose that these discrete sample points are used to interpolate and uniformly resample (in the (θ, y) space) the partial 3D shape of the object

$r_i(\theta, y)$ for the i-th view. For any two successive views, say i-th view and j-th view where $j = (i+1)\, mod\, 8$, assume that the object's surface overlaps partially. In the overlapping parts, assume that the resampling of r_i and r_j have been done for the same values of (θ, y). Let the position of the rotation axis (X_0, Z_0) be slightly different for different y. In this case, we can denote the position by $(X_0(y), Z_0(y))$. Now, we can improve the estimates for $(X_0(y), Z_0(y))$ by minimizing the squared distance between the same points in the successive views in the overlapping parts as follows:

$$\varepsilon(y) = \sum_{i=0}^{7} \sum_{\theta} (r_i(\theta, y) - r_j(\theta, y))^2 \tag{5}$$

where $j = (i+1)\, mod\, 8$. In the above equation, the summation over θ is done for those values of θ for which we have common resampled points on the overlapping surfaces from i-th and j-th views for $j = (i+1)\, mod\, 8$. The error measure $\varepsilon(y)$ is computed for various values of $(X_0(y), Z_0(y))$ near the initial estimates (X_0, Z_0). In the experiments, this was done at 1 mm intervals in the range $[X_0 - 50, X_0 + 50]$ and $[Z_0 - 50, Z_0 + 50]$. The interval for resampling in the θ space was 1 degree. The position where the error measure was a minimum was taken to be the correct position of the rotation axis.

In the integration method above, the rotation axis is assumed to be *inside* the object. If this assumption is not valid, then we need to modify our method. One alternative is to shift the origin of the object coordinate system for each cross-section of the object to some point inside the cross-section. For example, we can choose the perimeter centroid of the cross-section as the origin of the object coordinate system. This improvement will be incorporated in the next version of the system.

4. EXPERIMENTS

The earlier version of Stonybrook VIsion System (SVIS)[7,8] was almost completely rebuilt except for the digital camera used (Fig. 3). It includes a new PC (Pentium II, 450 MHz), a new fast digitizer (Matrox Meteor II , Standard), two computer controlled motors (MD-2 of Arrick Robotics)– one for moving the camera, and another for rotating a test object, and the Olympus DELTIS VC 1000 digital camera with a motorized lens which is controlled by the PC. This system was calibrated for Image Focus Analysis (IFA), Image Defocus Analysis (IDA), and Stereo Image Analysis (SIA).

SVIS was tested on three objects with a random color pattern pasted on them. Instead of pasting the random color pattern, one can project a similar color pattern using a slide/overhead projector. The purpose of the pattern is to introduce high contrast on the surface of the object to facilitate stereo matching. For each object, eight views at 45 degree intervals were used. The results are presented in Figs. 7-9. SVIS was also tested on the same objects using only 4 views taken at 90 degree intervals. The results were somewhat worse than that for eight views for our test objects.

Figure 7 shows the results for a sculpture of a face. It shows the focused image for the front view (a), partial 3D models for front and back views (b,c), wire frame plots of the side and top views of 3D shape (d,e), the complete 3D shape (f), and the complete 3D model with texture mapping on the face (g).

Figure 8 shows the results for an upright rectangular box with three different geometric shapes– a half-cylinder, a prism, and a cone– pasted on its 3 faces. A wire frame plot of the top view, the complete 3D shape, and the complete 3D model with texture mapping are shown.

Figure 9 shows the results for an upright cylinder with four different geometric shapes– a half-cylinder, a cone, a prism, and a cone-like polyhedron with a pentagonal base– pasted on its outer surface. A wire frame plot of the top view, the complete 3D shape, and the complete 3D model with texture mapping are shown.

A quantitative analysis of the accuracy of the results will be done in the future.

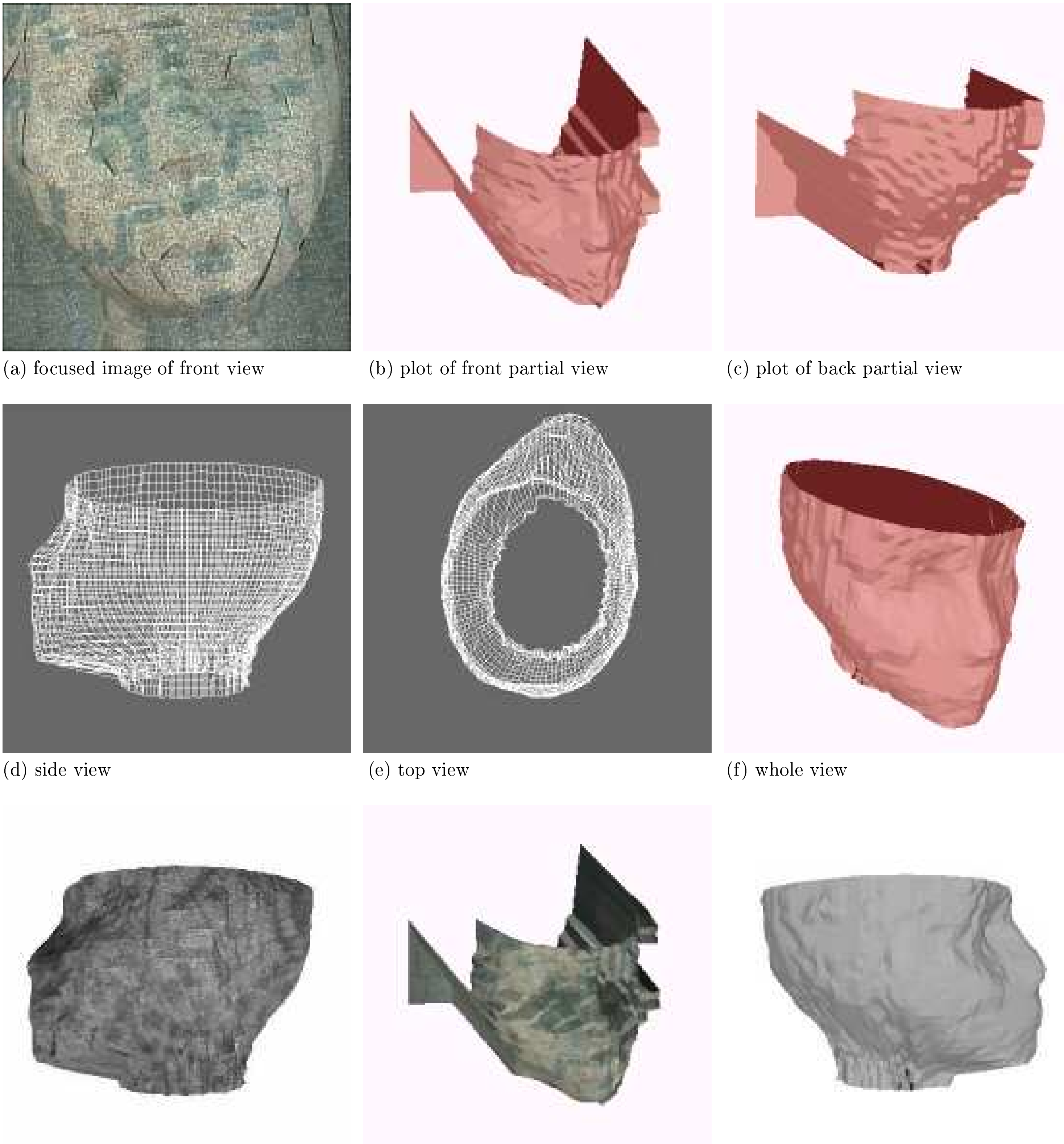
5. CONCLUSION

We have described a digital vision system for 3D model recovery using focus analysis and stereo matching. Partial 3D models are acquired from eight views of objects, and the partial models are integrated into a complete 3D model. Computational algorithms for stereo matching and partial model integration are presented. Experimental results are presented for 3 representative objects. In the future, this work will be extended to get more accurate results with fewer views. Also, we will extend this work to deal with more complicated objects than those considered here.

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

REFERENCES

1. R.M. Haralick and L.G. Shapiro, *Computer and Robot Vision*, Addison-Wesley Publishing Co., Inc, 1992.
2. B. K. P. Horn, *Robot Vision,* McGraw-Hill Book Company, 1986.
3. E. Krotkov, "Focusing", *International Journal of Computer Vision*, 1, 223-237, 1987.
4. M. Subbarao and T. S. Choi, "Accurate Recovery of Three-Dimensional Shape from Image Focus", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* March 1995, pp. 266-274.
5. M. Subbarao and Y. F. Liu, "Accurate Reconstruction of Three-Dimensional Shape and Focused Image from a Sequence of Noisy Defocused Images", *Proceedings of SPIE*, Vol. 2909, Nov. 1996, Boston, pp. 178-191.
6. M. Subbarao and G. Surya, "Depth from Defocus: A Spatial Domain Approach", *International Journal of Computer Vision*, 13, 3, pp. 271-294 (1994).
7. M. Subbarao, T. Yuan, and J. K. Tyan, "Integration of Defocus and Focus Aalysis with Stereo for 3D Shape Recovery", *Proceedings of SPIE*, Vol. 3204, pp. 11-23, Photonics East, Oct. 1997.
8. T. Yuan, and M. Subbarao, "Integration of multiple-baseline color stereo vision with focus and defocus analysis for 3D shape measurement", *Proceedings of SPIE*, Vol. 3520, pp. 44-51, Nov. 1998.

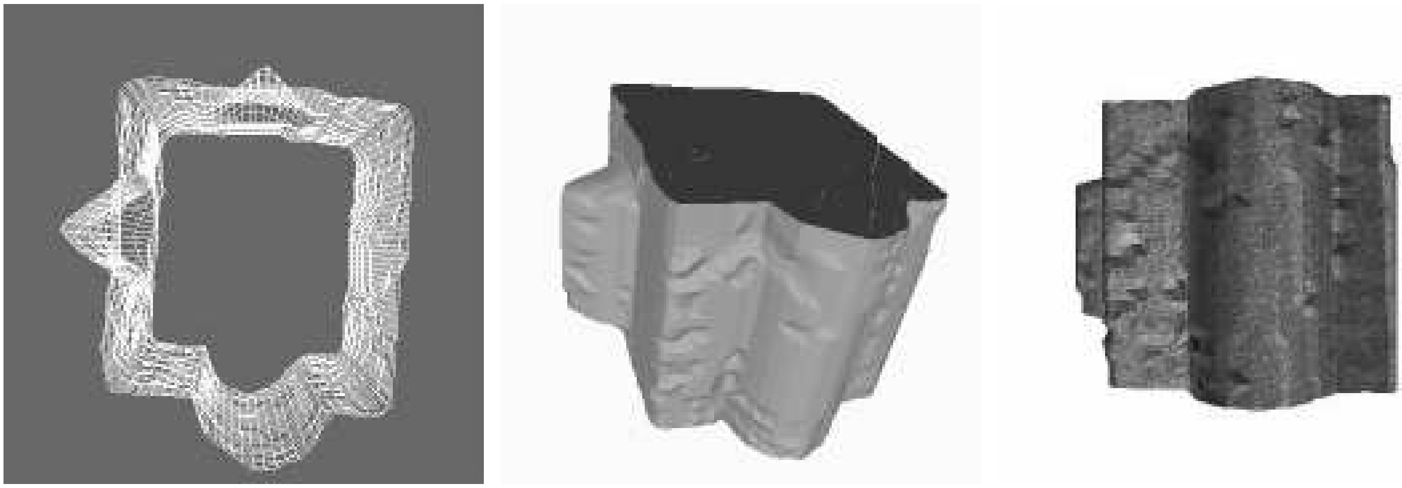


(a) focused image of front view (b) plot of front partial view (c) plot of back partial view

(d) side view (e) top view (f) whole view

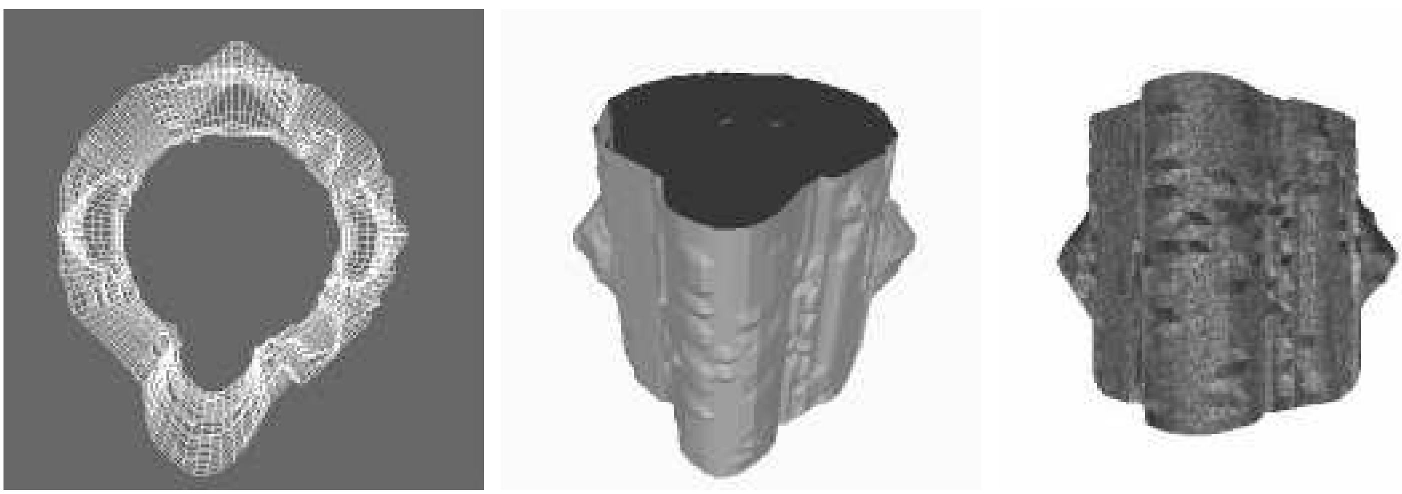
(g) side view with texture mapping (h) front view with texture mapping (i) side view without texture

Figure 7. Face object



(a) top view (b) whole view (c) with texture mapping

Figure 8. Box with half cylinder, prism and cone



(a) top view (b) whole view (c) with texture mapping

Figure 9. Cylinder with half cylinder, cone, prism and pentagon