Autonomous Robots 6, 89–103 (1999)
© 1999 Kluwer Academic Publishers. Manufactured in The Netherlands.

Human-Assisted Virtual Environment Modeling for Robots

J.G. WANG* AND Y.F. LI
Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong
meyfli@cityu.edu.hk

Abstract. In order to avoid the complex calculation and poor robustness in automatic visual modeling process, a man-machine interaction based stereo vision system is developed for modeling an unknown environment. The operator's knowledge about a scene is used as a guidance for modeling 3D environment. The modeling technique has advantage in terms of reliability and robustness over other automatic modeling approaches. The data points needed for modeling an objects are obtained through the intersection of lines, or calculation from equations of curve, derived via fitting from human guided edge detection. The modeling accuracy is ensured by using image feature extraction. A multi-viewpoint modeling approach has been developed in order to deal with occlusion problems.

Both accuracy and speed issues are addressed in this paper. The system implementation and some 3D measurements on real scene have been performed using cameras lenses of 16 mm and 8 mm with an accuracy 0.5 mm and 0.8 mm over the field of view, respectively. The virtual environment rendering based on the modeling data of real scenes with known model of mobile robot is given at the end of this paper.

Keywords: virtual reality, modeling, man-machine interaction

1. Introduction

Virtual Environment (VE) modeling has been a key problem in robotics (Ishiguro et al., 1995). In many robotic systems, e.g., telerobotic manipulator system, an operator utilizes 2D images from a remote camera to execute the task (Johnson et al., 1995). This limits the effectiveness of such teleoperated system since an image provides only 2D information. 3D information is of vital importance in many applications. Many researchers have studied the issues on how to build virtual environment using images taken from vision sensors while exploring the unknown environment, including automatic modeling (Chen and Trivedi, 1993) and semi-automatic modeling (Johnson et al., 1995), with minimum human interaction employed in the latter. Due to the complexity and robustness issues involved in automatic modeling, a new method has been studied to combine human and machine intelligence for environment modeling, where the use of virtual reality graphics environment provides an efficient method, such that human operators can use their knowledge about the geometric or physical attributes of objects in a real scene and guide the robot in modeling the environment and executing the task. An integrated robotic manipulator system using virtual reality concepts has been intensively studied in (Chen and Trivedi, 1993; Trivedi and Chen, 1993), where real-time simulation, visualization are utilized to create advanced, flexible and intelligent user interface. One of their planned activities of this project is to best match/build the virtual world model to the real world automatically, based on real and simulated sensory information. A virtual reality calibration technology is provided in (Bejczy et al., 1990; Kim, 1994; Kim et al., 1993; Kim, 1996) where an operator-interaction method was adopted to provide the correspondence information between 3D object model points of the robot arm and 2D camera image points, as well as the reliable correspondence data for object localization. Interactive model building was set as the planned work of their project. Interactive perception utilizes computer power for precision

**Current address*: Intelligent Machines Research Lab., Division of Control & Instrumentation, School of EEE, Nanyang Technological University, Singapore 639798.

measurement, and human perception for recognition, scene segmentation, and approximate location designation where reliable and efficient computer algorithms are unavailable (Backes et al., 1994). An interactive modeling system was proposed to allow modeling remote physical environment by two CCD cameras, voice command (Cooke and Stansfield, 1994; Miner and Stansfield, 1994). Edge information was used for stereo matching and triangulation to extract geometric and positional information of the object. The vision system was able to extract and model blocks and cylinders only. The block model was used as a bounding box for more complex objects. The system was limited to camera motion about Z axis only. A human-robot interface, which integrates real images, characters, diagrams, and voice, was adopted in (Nakashima et al., 1995). A model-based robot system, which executes a manipulation task, is presented in (Hasegawa et al., 1991). The 3D data is provided by laser pointer, but no occluded part processing was considered.

There are two distinct types of environments: known and unknown. The location and orientation of the objects in the first case can be reconstructed by using monocular vision with known models of objects (Michel et al., 1989). For the latter, the environment can be modeled with binocular stereo vision and multi-viewpoint observation strategy. In order to avoid the complex calculation of image feature matching and poor robustness of binocular stereo vision on unknown environment modeling, a man-machine interaction based stereo vision system is proposed in this paper. It is easy for human to recognize the objects and give some cues in real scene, but it is time consuming and complicated for a robot. In this system, interactive modeling is used to reconstruct an environment model through an operator giving the system necessary minimum cues about object attributes and features, matching methods and so on. The precision of the modeling method is ensured by human guided edge detection and line or curve fitting technique. The features needed for modeling an object are obtained by the intersection of the lines or by calculation of equation of the curve. A multi-viewpoint vision modeling scheme is proposed to deal with occlusion problems. First, local models of objects are built from different viewpoints. Then a global 3D model of an environment is reconstructed by matching and merging of the local models. When the environment is constructed, we render a virtual environment to allow an operator to see the environment from any viewpoint and to teleoperate robot to execute the task, e.g., grasp or part mating, more reliably. In this way, human intelligence can be effectively employed in executing robotic tasks. The system implementation and the experimental results are presented.

We present our man-machine interactive modeling in Section 2. Then the multi-viewpoint modeling for occlude situation is addressed in Section 3. Finally, some experiment results are given in Section 4, followed by some conclusions in Section 5.

2. Man-Machine Interaction-based Modeling

A stereo vision uses two cameras, the left and right. When calibrated, two transformation matrices, $[H_L]$ and $[H_R]$ between the cameras and a world frame (defined as W) are obtained respectively. The 3D coordinates of feature points in W, corresponding to the known image coordinate of feature points, can be calculated by using $[H_L]$ and $[H_R]$. Here, both $[H_L]$ and $[H_R]$ are $[3 \times 4]$ matrices.

Assume a 3D vector in W is represented by $[V_{3d}]$ and its correspondent 2D vector on image is represented by $[V_{2d}]$. Using a least-squares fitting algorithm, there exists

$$[H] = [V_{2d}][V_{3d}]^T[[V_{3d}][V_{3d}]^T]^{-1} \tag{1}$$

We can calculate the rotation and translation parameters of $[H]$, as well as the focal lengths and image center coordinates of the two cameras by decomposing matrix $[H]$.

Assuming $[H_L]$ and $[H_R]$ are available, we can then calculate the 3D coordinate $[X] = [x, y, z]$ of a feature point in W with its corresponding image coordinates $[x_a, y_a]$, $[x_b, y_b]$ on the two images

$$[A][X]^T = [B] \tag{2}$$

where $[A] = [a_{ij}], i = 1, \ldots, 3, j = 1 \cdots 4, [B] = [b_j], j = 1, \ldots, 4$, can be obtained from $[H_L]$ and $[H_R]$, and $[X]$ can be obtained

$$[X] = [[A]^T[A]]^{-1}[A]^T[B] \tag{3}$$

A major difficulty in stereo vision is the correspondence problem between the feature points in two images. Existing feature extraction and matching algorithms suffer from the poor robustness. However, a human operator can easily identify the objects in most scene images. The operator can prompt the vision system to locate and detect some object attributes or special corresponding features (such as edges and vertices of objects) in an image by man-machine interaction, so that the image coordinates of the features can be acquired and their 3D position in W calculated using Eq. (3).

Using the man-machine interaction paradigm, an operator can carry out the modeling task easily and robustly by utilizing VR techniques. In our work, we also developed a human guided feature extraction scheme in order to ensure the modeling accuracy. Since an object can be defined as a composite of some primitive models. An operator can recognize the attribute of object primitives and guide the vision system to find some correspondent feature points of the primitives using human guided edge detection in the image. Then a binocular stereo vision system can be used to construct the local models of objects directly. Although there may exist many kinds of objects in an environment, the modeling system through man-machine interaction can reconstruct the models of objects by merging all primitive models and eventually build whole environment model.

2.1. Primitive and Composite Models

In our test system, the primitive models consist of cuboid, sphere, cylinder, cone etc. An object, e.g., a table, can be represented by a composite of those primitives.

2.1.1. Cuboid. A cuboid can be reconstructed using its four vertices. In our system, the four vertices of a cuboid are given by an operator when he points out the corresponding points in the images which are defined as the feature points of the cuboid. For instance, there are four points for a cuboid, numbered as 1, 2, 3, 6, pointed by an operator on the left and right images, respectively. The cuboid will be determined in a world frame when the corresponding 3D coordinate (x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3) and (x_6, y_6, z_6) of the four vertices are calculated by the stereo vision algorithm, as shown in Fig. 1.

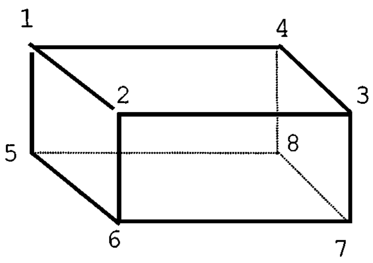


Figure 1. Definition of the feature points for modeling of a cuboid.

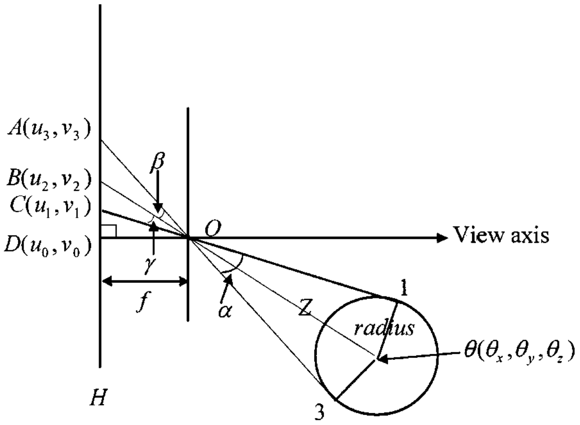


Figure 2. Modeling of a sphere.

2.1.2. Sphere. In general, the projective image of a sphere is an ellipse. We can fit the ellipse using at least five points on the ellipse. Then the coordinates of image points (u_1, v_1), (u_3, v_3), which are the two vertices of the major axis of the ellipse as shown in Fig. 2, can be calculated. The angle of $\angle AOC$, defined by the points (u_1, v_1), (u_3, v_3), and the center of camera (i.e., point $O(c_x, c_y, c_z)$ in Fig. 2), is equally divided by the projective line $(B\theta)$, giving $\beta = \gamma = \alpha/2$ as shown in Fig. 2. We can then calculate the image coordinates (u_2, v_2) of the sphere center θ. As shown in Fig. 2, in $\triangle AOC$, we have:

$$AB/BC = AO/CO.$$

Then we arrive at:

$$\begin{aligned} u_2 &= (u_1 + r * u_3)/(1 + r), \\ v_2 &= (v_1 + r * v_3)/(1 + r) \end{aligned} \tag{4}$$

where $r = ((u_1 - u_0)^2/k_1^2 + (v_1 - v_0)/k_2^2 + 1)^{1/2}/((u_3 - u_0)^2/k_1^2 + (v_3 - v_0)^2/k_2^2 + 1)^{1/2}$ and $k_1 = f/u_r, k_2 = f/v_r, u_r$ is the length of a pixel in u direction, v_r is the length of a pixel in v direction on the image, and f is the focal length of the camera. The 3D coordinate $(\theta_x, \theta_y, \theta_z)$ of the sphere center can be obtained by the binocular stereo vision algorithm when the image coordinates of the sphere center are calculated using the above equations in the left and right images, respectively. If the 3D coordinate of the camera center is (c_x, c_y, c_z), obtained from camera calibration, then the radius of the sphere will be

$$\text{radius} = Z * \sin(\alpha/2) \tag{5}$$

where

$$Z = \sqrt{(c_x - \theta_x)^2 + (c_y - \theta_y)^2 + (c_z - \theta_z)^2}$$

α is calculated using the cosine law within $\triangle ACO$ using points $A(u_1, v_1)$, $C(u_3, v_3)$, $D(u_0, v_0)$. This gives

$$AC^2 = AO^2 + CO^2 - 2 * AO * CO * \cos(\alpha)$$
$$AC^2 = AD^2 + DO^2$$
$$CO^2 = CD^2 + DO^2$$

Then we have

$$\alpha = \cos^{-1}\left(\left(r_1^2 + r_2^2 - r_3^2\right)\big/(2 * r_1 * r_2)\right) \qquad (6)$$

where

$$r_1 = \left((u_1 - u_0)^2\big/k_1^2 + (v_1 - v_0)\big/k_2^2 + 1\right)^{1/2}$$
$$r_2 = \left((u_2 - u_0)^2\big/k_1^2 + (v_2 - v_0)\big/k_2^2 + 1\right)^{1/2}$$
$$r_3 = \left((u_2 - u_1)^2\big/k_1^2 + (v_2 - v_1)\big/k_2^2\right)^{1/2}$$

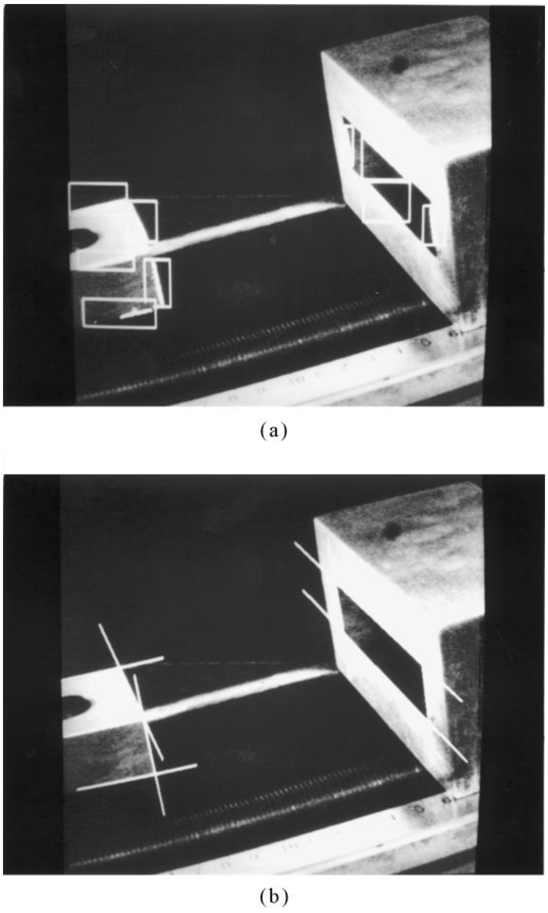
A sphere is thus determined in 3D space when its center and radius are given.

2.1.3. Other Primitives. Other primitives can also be constructed in a way similar to the above. For example, since both the top and bottom of a cylinder are circles of the same radius and the projective image of the two circles are elliptic, we can fit the two ellipses using at least five points on them. Then, the pose and position of the cylinder can be determined using quadratic curve based on stereo vision. A circular cone can be processed similarly, with a vertex needed. Other polygonal cones can also be modeled by their vertices. For example, a triangular pyramid can be determined by four vertices, i.e., one top vertex and three bottom vertices.

2.1.4. Composite Objects. Composite objects can be modeled by using the primitive models. For example, a table is composed of some cuboids. It can be built using a top cuboid and a leg cuboid (assume the legs are symmetric in their geometry). The man-machine interactions will be used in all the modeling exercises. Other objects such as door, window, wall can also be built in a similar way. For example, a wall is formed by its corner points.

2.2. Operator Guided Feature Extraction

In order to improve the precision and reliability in the modeling, an operator guided edge detection method was adopted in our work. All of the edges needed for modeling an object can be extracted one-by-one through human-machine interactions. All the user needs to do is to draw a rectangle that encompass image edge using mouse and to give a threshold. The edge detection then will be done based on an detection operator and the threshold specified. The edge detection results will be fitted as a line or curve using least-squares fitting algorithm. The vertexes of the object are found through the intersection of the corresponding lines. Different edge detectors, such as Roberts, Sobel, Laplacian, Kirsch, and threshold can be used as selected by the user. The region following method has been used before in the fitting operation in order to reduce the effect of noise. Figure 3 shows the feature


(a)

(b)

Figure 3. Human guided feature extraction.

extraction in a part mating experiment. Figure 3(a) shows the result of edges detection and Fig. 3(b) shows the line fitting result and vertexes location through the intersection of the lines. In general, a vertex can be found by the intersection of two or more lines.

For other objects, similar approach can be adopted in using operator guidance. For example, a sphere or cylinder feature extraction can be achieved by edge detection and corresponding ellipse fitting, i.e., vertexes of major axis.

3. Multi-Viewpoint Modeling

In general, a global description of objects cannot be reconstructed from only one viewpoint due to occlusion or limited field of view in a real situation. Therefore, a multi-viewpoint modeling strategy needs to be developed. In this paper, we developed a multi-viewpoint modeling strategy that uses controlled view points. In this way, the local models of objects or environment are built by different viewpoints separately. With these local models, the global model of an object or environment can be derived by merging the local models.

The scene images observed from the different viewpoints are changed as the camera moves around. The feature representations also change corresponding to the different frames when the viewpoint is changed. So we must know the pose transformation of the vision system between two viewpoints. Here, a special method is used in obtaining the transformation (M) by using at least two common feature vectors of some objects which are taken as images from different viewpoint (Dong et al., 1997; McCarthy, 1990). The common feature vectors are pointed by man-machine interaction.

The scheme mentioned above is illustrated in Fig. 4, where A and B represent two viewpoints 1 and 2 respectively, W represents an object frame in the world space. C and C' represent the coordinate relationships between A and W, B and W respectively, and M represents the pose transformation of the vision system between A and B. When the vision system moves from viewpoint A to viewpoint B, the object will have a transformation M^{-1}. If we can determine M, then the two local models of the object which are modeled from two viewpoints separately, can be merged to generate a global model of the object.

Let W and W' represent the 3D poses of the object in frames A and B respectively, then we have

$$C' = M^{-1}C \tag{7}$$

$$W = M^{-1}W' \tag{8}$$

In determining M, we solve for its rotation and translation separately. First, a space vector defined by two special points on an object is obtained using the stereo vision system observing from two viewpoints. With the property of invariant 3D coordinate of a rigid body, the rotational relationship between W and W' can be given (from Rodrigues equation) by

$$(W - W') = U(W + W') \tag{9}$$

and

$$R = [I + U][I - U]^{-1}$$

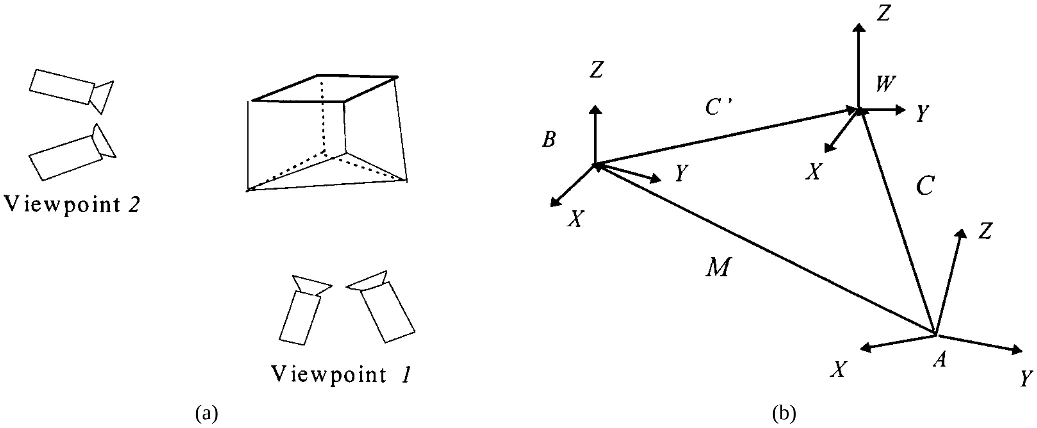


Figure 4. (a) Procedure of multi-viewpoint modeling and (b) transformation of multi-viewpoint modeling.

To calculate U, there will be at least two such vectors which could be obtained from three feature points on the object. These feature points could be taken by stereo vision with man-machine interaction. R will be obtained when U is calculated and translation of M is calculated by $T = W - [R]W'$. Here

$$M = [R \quad T] \tag{10}$$

where

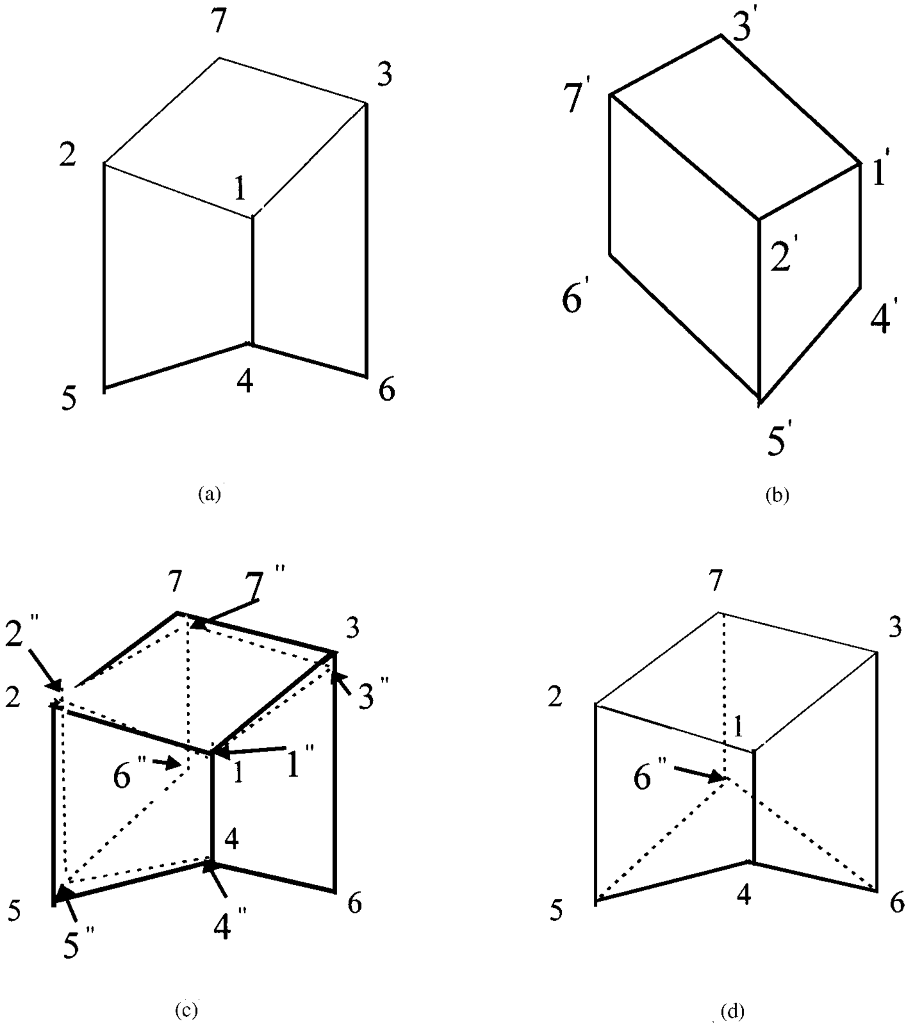
$$R = \begin{bmatrix} R_{00} & R_{01} & R_{02} \\ R_{10} & R_{11} & R_{12} \\ R_{20} & R_{21} & R_{22} \end{bmatrix}, \quad T = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}$$

The 3D modeling data from viewpoint B can be transformed to viewpoint A using M,

$$W'' = [R]W' + [T] \tag{11}$$

Using W'', the whole model of the object can be obtained through the complementary 3D data that can be calculated from viewpoint 2 but cannot be calculated from viewpoint 1.

An example of a simple object modeling is shown in Fig. 5, where Fig. 5(a) and (b) represent the local models of an object that were constructed from different viewpoints A and B, respectively. Figure 5(c) represents a translated 3D model of Fig. 5(b), which is constructed from viewpoint 2 into viewpoint 1 (dot line) and overlaid with Fig. 5(a). After the edge detection



(a)

(b)

(c)

(d)

Figure 5. Matching of multi-viewpoint modeling (1 and 1′, 2 and 2′, 3 and 3′, 4 and 4′ are common points).

and feature extraction, the operator prompts the vision system for the common points: point 1 and point 1′, 2 and 2′, 3 and 3′, 4 and 4′ in the images taken from two viewpoints. The translations between these points are represented as the vectors we need. Then, M can be calculated as above followed by merging of the local models in Fig. 5(a) and (b) with M. The global model can be reconstructed when the local models are enough to merge the object completely. For example, point 6″ in Fig. 5(c), transformed from point 6′ in Fig. 5(b) using M, will be added to the 3D local model of viewpoint 1 in order to obtain the whole model of the object shown in Fig. 5(d).

The precision of the procedure can be given in two ways. Firstly, it can be represented by the distance between every two common points in whole model, such as points 1 and 1″, 2 and 2″, 3 and 3″, 4 and 4″, 5 and 5″, 7 and 7″ in Fig. 5(c). Secondly, it can be represented by the difference between measured and real size of the object. In this paper, a non-regular cube has been modeled using the multi-viewpoint method. The precision obtained using above two ways are both less than 0.5 mm in our experiments. The multi-viewpoint algorithm we have developed can be extend easily to the situation that over two viewpoints are needed for modeling a whole object/scene.

4. Man-Machine Interaction and Experimental Results

4.1. *Configuration and Calibration of the Stereo System*

As an application, the man-machine interaction based stereo vision system as an interface for a teleoperated mobile robot is built using a PC-based computer system. A hierarchical structure in the virtual environment database and a special description language of virtual environment are also developed on an SGI workstation. The results in modeling an environment are transferred from the PC to an SGI workstation via local internet or to another PC via RS-232 serial port. The system architecture is shown in Fig. 6.

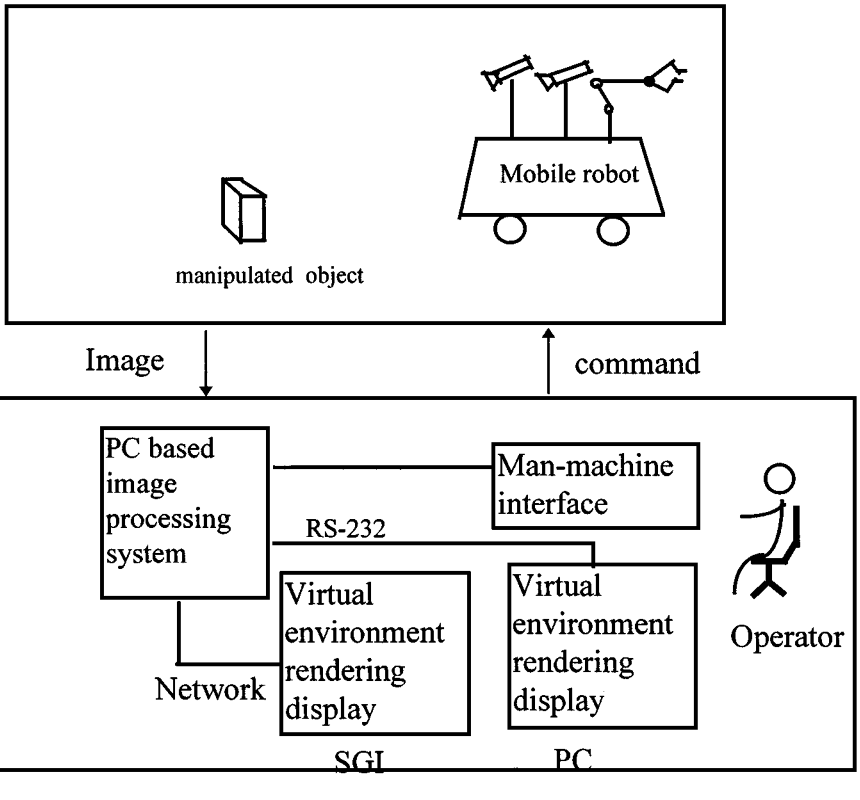

Mobile robot
manipulated object
Image
command
PC based image processing system
Man-machine interface
RS-232
Virtual environment rendering display
Virtual environment rendering display
Network
Operator
SGI
PC

Figure 6. The system architecture.

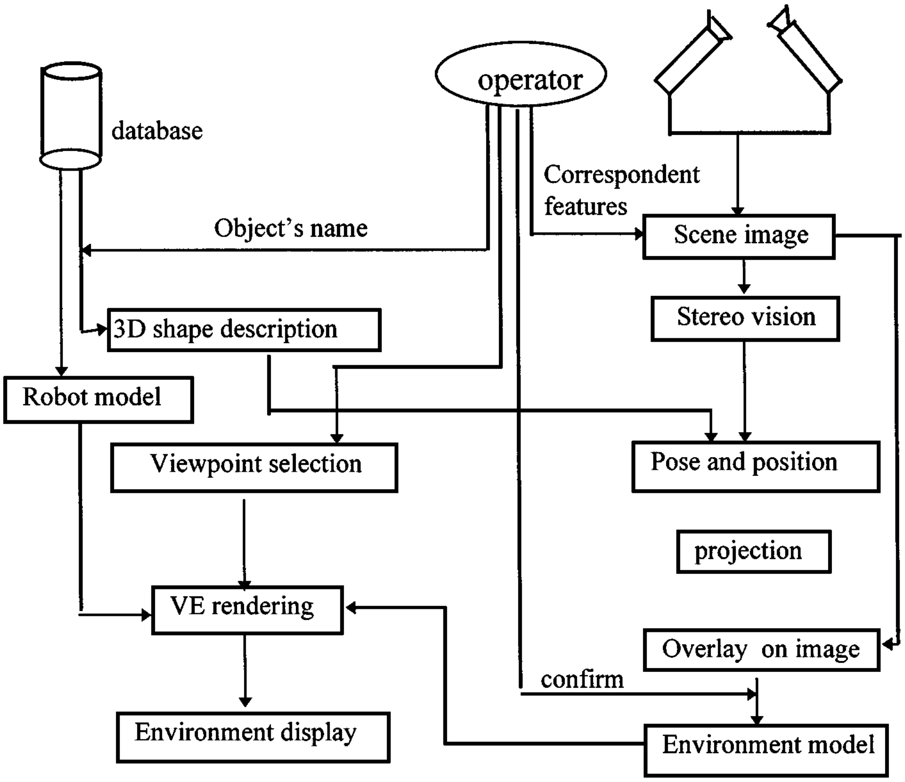


Figure 7. The modeling system flowchart based on man-machine interaction.

The operator can observe and recognize objects in a real scene on the monitor. He can also prompt the system for some object's name and guide their edge detection in finding correspondent feature points needed for modeling the object, on the screen through the interface. The 3D data of the features and transformations of the vision system between different viewpoints are calculated automatically. Then, the models of objects are overlaid to the original images to verify the modeling results. Finally, the environment will be reconstructed with these models and their locations in a reference frame. It will be transmitted to SGI workstation when the operator confirms it. The known models of mobile robots and the manipulators are stored in the model database so that the global scene model can be generated in virtual environment automatically. The operator can observe the pose relationship between the robot and the obstacles or between the manipulator and the object from any viewpoints. In the virtual environment, the system offers a great help to an operator working in an unknown environment. This will strengthen the efficiency and reliability of the teleoperation. The schematic representation of the virtual environment modeling is shown in Fig. 7.

Figure 8 shows our human-machine interaction modeling system. Two *MINTRON MTV-1881EX* CCD cameras are used. The two cameras are connected to the red and green channel of a color image grabber, respectively. The internal synchronization is used, i.e., the two cameras are synchronized by the same output signal from the image grabber card. A plate marked with black squares is mounted on a one-dimension

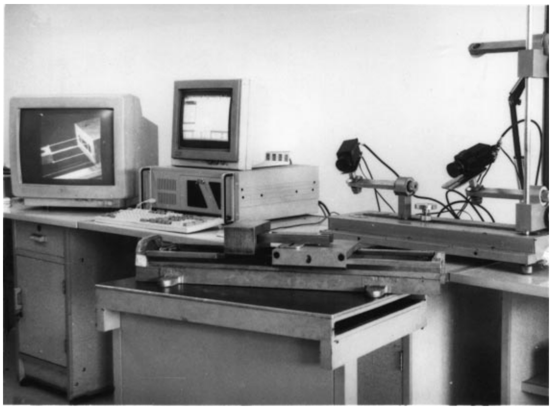


Figure 8. Human-assistance environment modeling system.

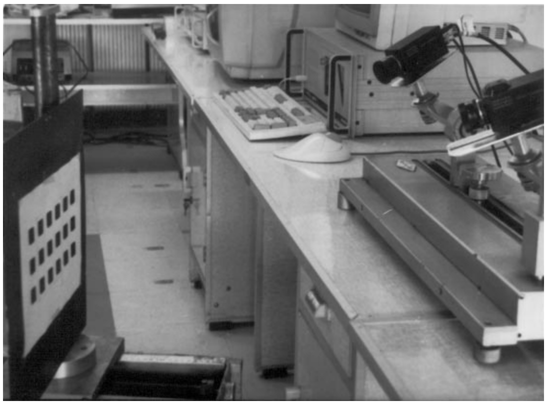


Figure 9. Experiment setup for cameras calibration.

movable calibration apparatus is used to calibrate the cameras, as is shown in Fig. 9. In order to improve the calibration precision, the vertexes of the squares marks are detected using human guided edge detection method as described in Section 2.1. The image coordinates of these vertexes and their corresponding 3D coordinates are used to calibrate the cameras using pseudo-inverse fitting. A total of 99 points are used in calibrating the two cameras. 9 points can be captured from one position and 11 steps each with a 10 mm step size are used in the calibration. Two AVENIR CCTV 16 mm lens cameras are calibrated. The transformation matrices derived are as follows:

$$[H_L] = \begin{bmatrix} 39.126310 & 8.214318 & 16.881193 & -100.456096 \\ -3.622237 & 43.831757 & -12.297145 & 331.765558 \\ -0.006779 & 0.006904 & 0.019769 & 1.000000 \end{bmatrix}$$

$$[H_R] = \begin{bmatrix} 47.205679 & -2.942710 & -11.715263 & -310.788084 \\ 0.584861 & 49.994324 & -13.079590 & 324.279067 \\ 0.009651 & 0.007713 & 0.021737 & 1.000000 \end{bmatrix}$$

To increase the precision of the stereo vision system, the 3D coordinates of another 99 points and 54 points are calculated using the corresponding image coordinates from the left and right images for lenses of 16 mm and 8 mm, respectively. Then they are compared with the known 3D coordinates. The errors of the stereo vision system with the calibrated cameras of 16 mm and 8 mm lenses are shown in Fig. 10(a) and (b), respectively. The errors are found to be less than 0.6 mm and 0.8 mm respectively over the field of view.

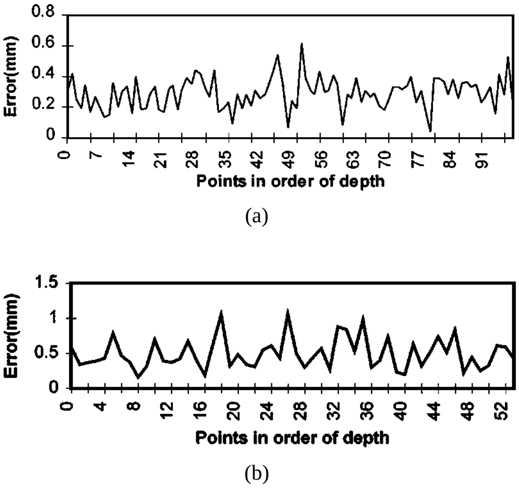


Figure 10. The errors of the vision system with (a) 16 mm and (b) 8 mm lens camera.

4.2. *Experiment Results*

Using the stereo vision system, we conducted several 3D experiments including object/environment modeling, multi-viewpoint modeling and vision guided part mating.

4.2.1. Object/Environment Modeling Experiment. Figure 11 shows an environment modeling example, where the environment consist of a desk, a sphere, a cuboid (book) and a cylinder. The necessary points for modeling these objects are detected. The desk, book and sphere can be modeled based on the extracted

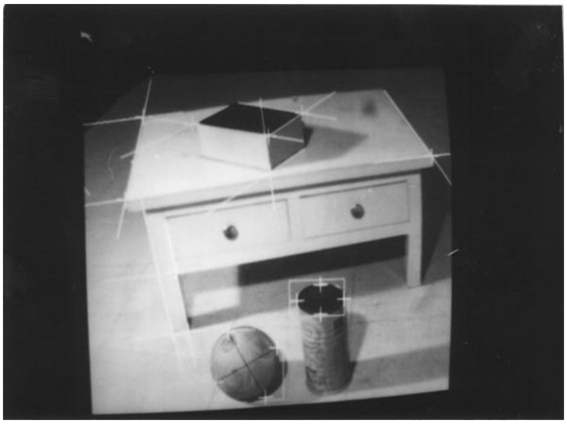


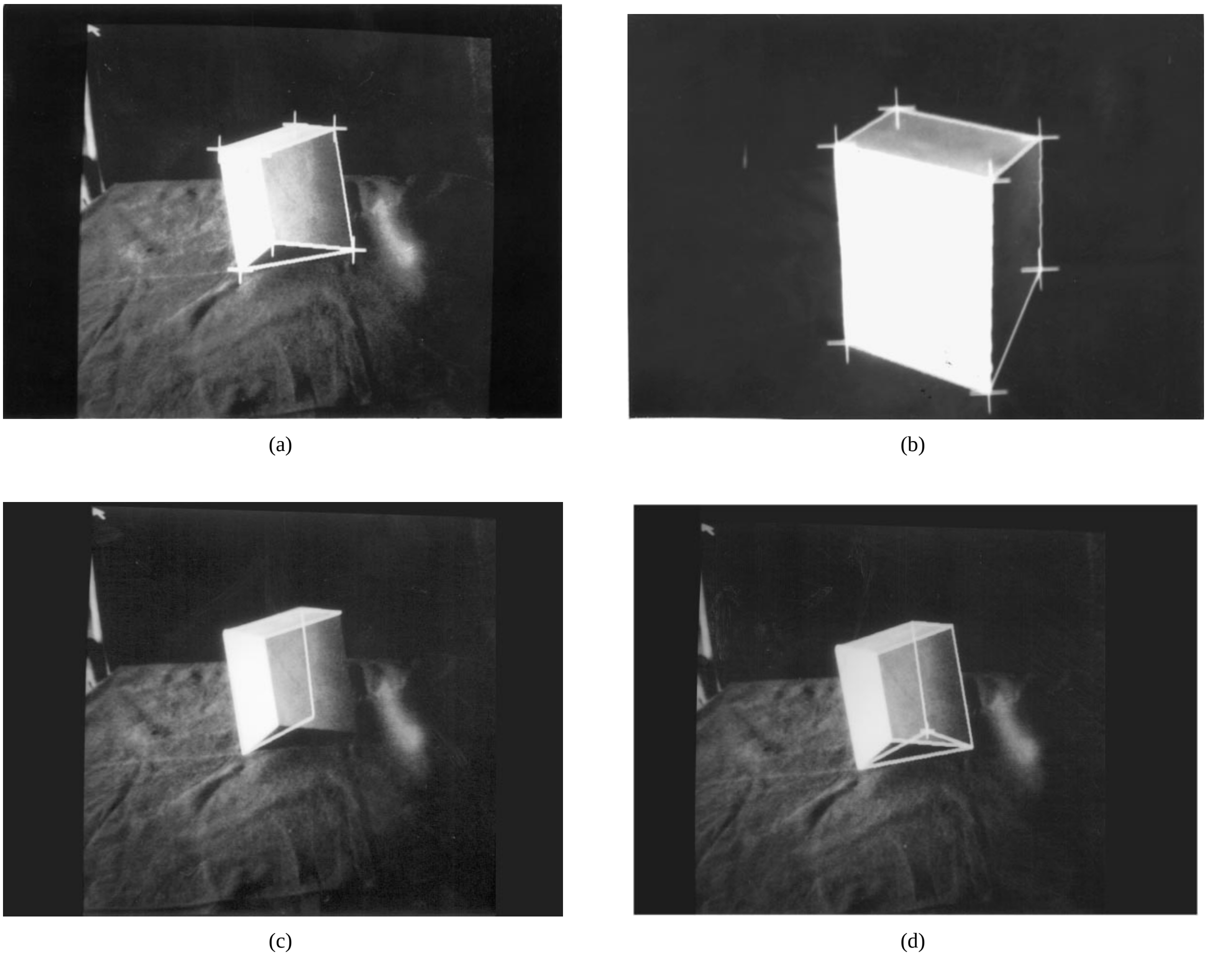
Figure 11. Object/environment modeling experiment.

features required as described in Section 2. The cylinder can be modeled using curve-based stereo vision method based on the extracted ellipse feature. The modeling accuracy turned out to be satisfactory. For example, the model of a sphere (with a radius of 1.885 cm) is tested using two cameras with the same lens 8 mm in two positions. After edge detection and ellipse fitting, the image coordinates (u_1, v_1), (u_3, v_3), which are the two vertices of the major axis of the ellipse shown in Section 2, can be calculated from the ellipse equation. In position 1, (285.73, 252.38), (407.35, 244.17) in the left image, (184.13, 254.25), (308.30, 248.63) in the right image. In position 2, (376.08, 80.45), (472.81, 108.06) in the left image, (60.87, 86.40), (159.62, 122.28) in the right image. The 3D coordinate of the center and the radius of the sphere are given in Table 1.

Table 1. The modeling result of a sphere.

	Measuring (cm)				Real (cm)	
Position	θ_x	θ_y	θ_z	Radius	Radius	Error of radius
1	5.79	5.43	17.46	1.869526	1.885	0.82%
2	5.57	5.09	31.09	1.871000	1.885	0.74%

4.2.2. Multi-Viewpoint Modeling. The multi-viewpoint algorithm we have developed is tested using a irregular cube. The experimental results are shown in Fig. 12. The modeled result in the first and second viewpoints are shown in Fig. 12(a) and (b) respectively. The projection from the second viewpoint to the first viewpoint using calculated rotation and translation matrix is shown in Fig. 12(c). The final modeling result of the irregular cube is shown in Fig. 12(d).



(a)

(b)

(c)

(d)

Figure 12. Modeling of a irregular cube using multi-viewpoint.

The rotation and translate transformation derived are as follows:

$$[R] = \begin{bmatrix} 0.234953 & 0.030250 & 0.971536 \\ -0.008533 & 0.999541 & -0.029057 \\ -0.971969 & -0.001468 & 0.235104 \end{bmatrix}$$

$$[T] = [-5.481626 \quad 0.455810 \quad 21.924350]$$

The distances of the six common vertexes are considered as errors as indicated in Section 3. The errors are given as follows, in the order of the vertex numbers given in Fig. 5:

Vertex	Error (mm)
1	0.000000
2	0.508750
3	0.259458
4	0.124624
5	0.174559
6	0.337252

It can be seen that the multi-viewpoint algorithm performed quite well with the errors smaller than 0.6 mm.

4.2.3. Virtual Environment Rendering. Using the above modeling method, we have modeled a scene captured by the vision system mounted on a mobile robot. The known models of mobile robot and the manipulator are stored in the model database so that the global scene model can be generated in virtual environment automatically. A virtual environment including the mobile robot and modeled 3D world is shown in Fig. 13. A rotation of about 90° is use between the viewpoints of the real vision system mounted on the mobile robot to observe the objects in the scene.

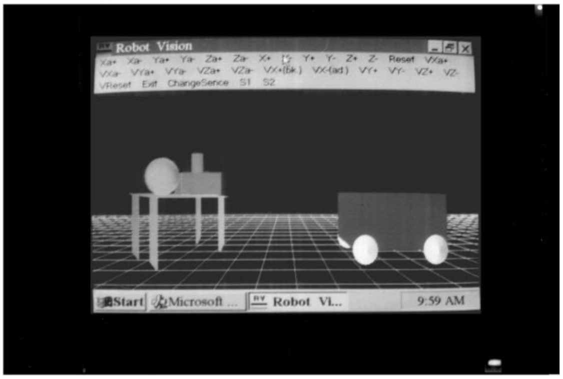

Robot Vision
Xa+ Xa- Ya+ Ya- Za+ Za- X+ X- Y+ Y- Z+ Z- Reset VXa+
VXa- VYa+ VYa- VZa+ VZa- VX+(bk.) VX-(ad.) VY+ VY- VZ+ VZ-
VReset Exit ChangeSence S1 S2
Start
Microsoft ...
Robot Vi...
9:59 AM

Figure 13. Virtual environment rendering.

4.2.4. Vision Guided Part Mating. Figure 14 illustrates the part mating experiment we conducted. Part 1 (left) is required to be inserted into part 2 (right). Here, we model part 1 with three points 1, 2, 3, and part 2 with another three points 4, 5, 6. These points can be calculated from the feature detection and stereo vision as described above. Then we can calculate the lines normal to the two planes π_1 and π_2, determined by points 1, 2, 3 and points 4, 5, 6, respectively. The angle between the two normal lines needs to be small enough so that the two planes are approximately parallel. Then the points 1, 2 and 3 are projected onto the plane π_2 (determined by $1'$, $2'$ and $3'$ as seen in Fig. 11) along the normal line vector of plane π_1. Part 1 can be inserted into part 2 if the following conditions are satisfied:

(1) The angle between line $1'$, $2'$ and line 4, 5 is small enough, i.e., line $1'$, $2'$ is approximately parallel to line 4, 5.
(2) Point $1'$ and $2'$ lie between line 4, 7 and line 5, 6.
(3) Point $2'$ and $3'$ lie between line 4, 5 and line 6, 7.

Some intermediate steps in carrying out the vision guided part mating task are shown in Fig. 15. Two cameras both with a 16 mm lens are used in the experiment. Part 1 is designed with a width of 50.2 mm and height of 19.9 mm. Part 2 is designed with the width of 102.6 mm and height of 20.3 mm respectively. The test result is given in Table 2.

In this table, width-1, height-1 and width-2, height-2 represent the width and length of part 1 and part 2 respectively. The *distance* in the table represents the distance between the two planes. Angle-1 represents the angle between the two normal lines of the two planes in Fig. 13. Angle-2 represents the angle between line 3, 4 and line $1'$, $2'$. The three steps brought part 1 into part 2 without further adjustment. In the three positions, the errors in the distance are calculated using the values of the real distance and the calculated distance via modeling approach. The real distance between the two planes of the parts are obtained from a ruler fixed on the movable equipment. These distances are 9.1 cm, 8.1 cm and 7.1 cm, respectively in the three steps. Figure 16 shows the results in another part mating experiment, with the measurement data listed in Table 3. This experiment shows the vision guided insertion when there exist a larger error in initial position of the peg which needs to be adjusted using the vision guidance.

Table 2. Some results in the part mating task.

Position	Width-2 (mm)	Height-2 (mm)	Width-1 (mm)	Height-1 (mm)	Distance (cm)	Angle-1 (degree)	Angle-2 (degree)	Error of distance
1	102.13658	20.09293	50.79605	19.77494	9.121807	0.862527	0.565099	0.241%
2	102.36857	20.05565	50.36557	19.53886	8.129533	0.807698	0.727662	0.35%
3	102.14636	20.08093	50.64573	19.67485	7.127684	0.813467	0.584093	0.39%

Table 3. Results of three steps of another part mating task.

Position	Width-2 (mm)	Height-2 (mm)	Width-1 (mm)	Height-1 (mm)	Distance (cm)	Angle-1 (degree)	Angle-2 (degree)
1	102.51714	20.28692	50.13752	19.04528	5.016932	5.022622	0.434312
2	102.65296	20.47162	50.87478	19.52097	4.090957	2.047162	0.161952
3	102.151714	20.28692	50.3958	19.91999	0.797795	1.633507	0.501241

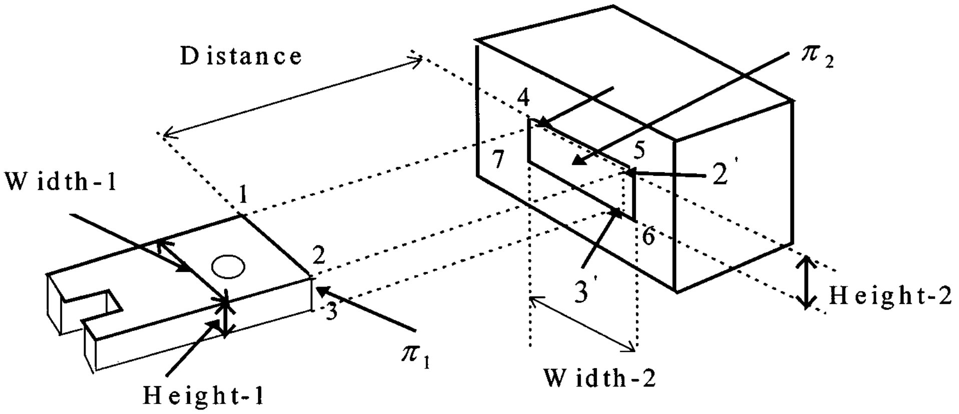
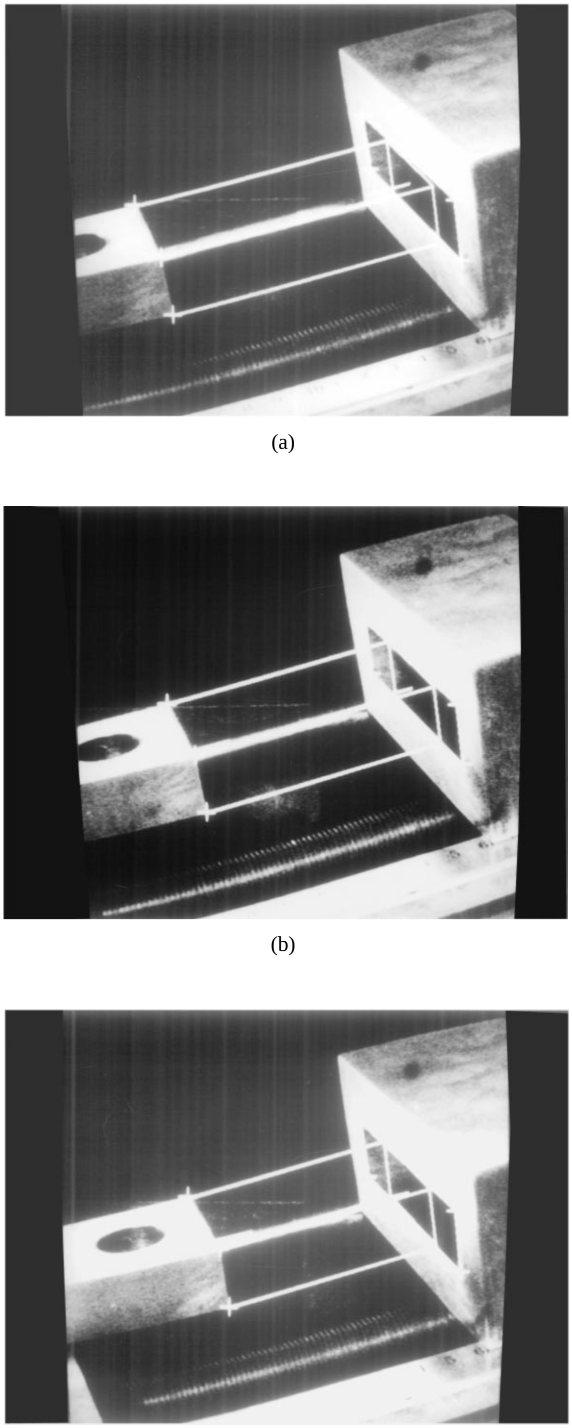


Figure 14. The vision guided part mating experiment.

In position 1, the part 1 needs to be tilted up by an angle of 5 degree and lifted up 1.2 cm based on the measured result. In position 2, part 1 is needs to be lifted up 0.1 cm. In position 3, part 1 has been inserted into part 2 correctly. In order to measure the mating accuracy, we remove part 2 while keeping part 1 in position 3, i.e., the position when part 1 is inserted in part 2, and then model part 1. Next using the model of part 2 derived previously, the gaps between part 1 and part 2 at position 3 are calculated and they are found to be 0.72224 mm and 0.34293 mm for the upper and lower planes respectively. The modeling result of part 1 is projected and overlaid on the previous case shown in Fig. 16(c). From measurements of the real parts, we know that the real gap between part 1 and part 2 is 0.4 mm. Therefore, we conclude that our vision system can execute this kind of part mating task when a 0.6 mm gap or more exists between part 1 and part 2. When a higher precision is required with a smaller tolerance given between the parts, force sensing needs to be employed to guide the mating process.

In the experiment here, the human-machine interactions (choosing the rectangles for the edge detection) constitute the most time consuming operation. The speed is measured in our test, where a *Pentium* PC with 16MB memory is used. Using the human-machine interaction, a vertex can be detected within 8 s in our experiment. The part mating experiment here should include the detection of 12 vertexes (6 vertexes on the left and right images respectively). Therefore, the time taken in such a modeling is about 1.5 min.

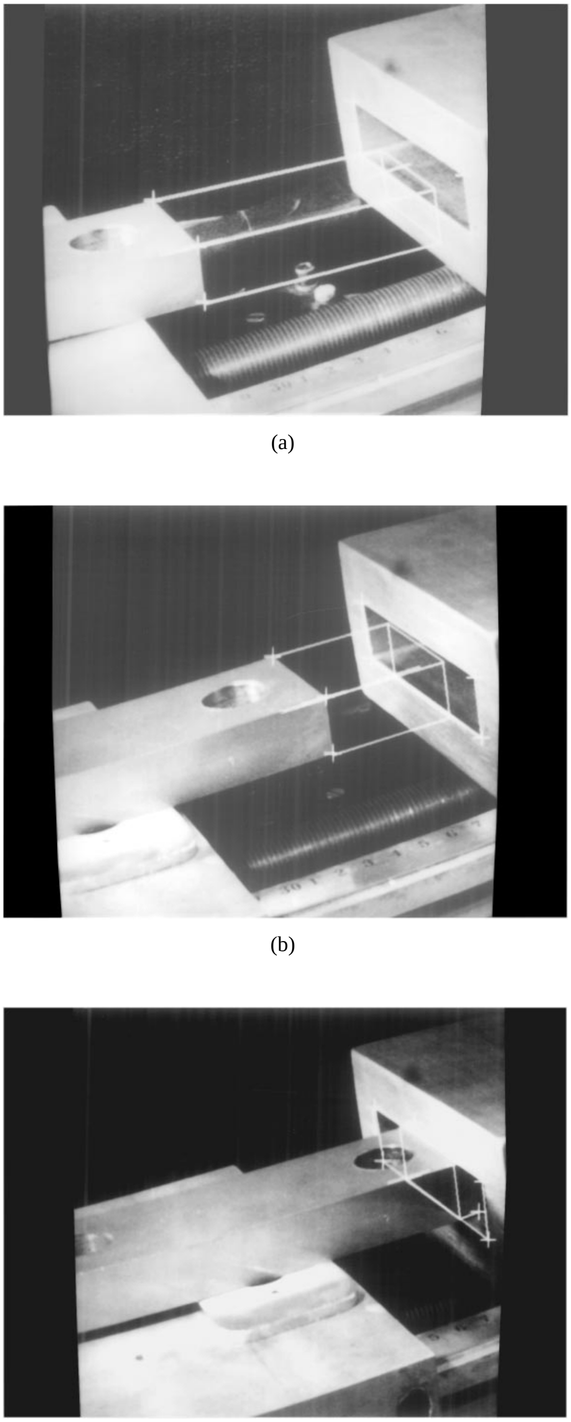


(a)

(b)

(c)

Figure 15. Some steps in a part mating experiment: (a) Position 1, (b) Position 2 and (c) Position 3.



(a)

(b)

(c)

Figure 16. Some steps of another part mating experiment: (a) Position 1, (b) Position 2 and (c) Position 3.

5. Conclusions

A virtual environment modeling system for robotic applications has been developed. A man-machine interaction and a multi-viewpoint observation strategy using a binocular stereo vision system have been studied to build an unknown environment model. The human guided image feature extraction and multi-viewpoint approach are theoretically and experimentally proven be viable for the environment modeling. The complicated calculation and poor robustness in automatic visual modeling are overcome due to the use of human knowledge about the scene. The efficiency and accuracy is measured in experiments. A 3D environment can be reconstructed based on the acquired models from stereo vision and the known models of a robot and manipulator on a PC or SGI workstation. Operators can observe the real environment on the screen and operate in the virtual environment from any viewpoints on the virtual reality system. By using human's intelligence, the system can greatly improve the teleoperation of a mobile robot working in an inaccessible environment. Future work is underway in investigating new camera calibration methods to further improve the accuracy of the modeling.

Acknowledgments

This work received partial support from the Research Grants Council of Hong Kong through grant number 9040368. The authors are grateful for the support and help of their colleagues Y. Hao, Z. Dong and Professor X. Xu at CAS.

References

Aubry, S. and Hayward, V. 1995. Three-dimensional model construction from multiple sensor viewpoint. In *Proc. IEEE 1995 Conference on Robotics and Automation*, pp. 2054–2059.

Backes, P.G., Beahan, J., Long, M.K., Steele, R.D., Bon, B., and Zimmerman, W. 1994. A prototype ground-remote telerobot control system. *Robotica*, 12:481–490.

Bejczy, A.K., Kim, W.S., and Venema, S.C. 1990. The phantom robot: Predicative displays for teleoperation with time delay. In *Proc. 1990 IEEE International Conference on Robotics and Automation*, pp. 546–551.

Chen, C. and Trivedi, M.M. 1993. SAVIC: A simulation, visualization and interactive control environment for mobile robots. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(1):123–144.

Cooke, C. and Stansfield, S. 1994. Interactive graphical model building using telepresence and virtual reality. In *Proc. 1994 IEEE International Conference on Intelligent Robotics and Systems*, pp. 1436–1440.

Dong, Z.L., Hao, Y.M., and Xu, X.P. 1997. LW-l measurement system of robot's performance. *ROBOT* (in Chinese), 19(5):35–41.

Fua, P. and Leclerc, Y.G. 1995. Object-centered surface recognition: Combining multi-image stereo and shading. *International Journal of Computer Vision*, 16:35–56.

Gagalowicz, A. 1995. Tools for advanced telepresence systems. *Comput & Graphics*, 19(1):73–81.

Ganapathy, S. 1984. Decomposition of transformation matrices for robot vision. In *Proc. IEEE Conf. on Robotics*, Atlanta, GA, pp. 130–139.

Gros, P. 1995. Matching and clustering: Two steps toward automatic object modeling in computer vision. *The International Journal of Robotics Research*, 14(6):633–642.

Hasegawa, T., Suehiro, T., and Takase, K. 1991. A robot system for unstructured environments based on an environment model and manipulation skills. In *Proc. of the 1991 IEEE International Conference on Robotics and Automation*, Sacramento, CA, pp. 916–923.

Ishiguro, H., Takeshi, M., and Takahiro, M. 1995. Building environment models of man-made environments by panoramic sensing. *Advanced Robotics*, 9(4):399–416.

Johnson, A.P., Hoffman, R., Hebert, M., and Osborn, J. 1995. 3D object modeling and recognition for telerobotic manipulation. *1995 IEEE Conference on Intelligent Robotics and Systems*, pp. 103–110.

Kim, W.S. 1994. Virtual reality calibration for telerobotic servicing. In *Proc. IEEE Conference on Robotics and Automation*, 4:2769–2775.

Kim, W.S. 1996. Virtual reality calibration and preview/predictive displays for telerobotics. *Presence*, 5(2):173–190.

Kim, W.S., Schenker, P.S., Bejczy, A.K., and Hayati, S. 1993. Advanced graphics interfaces for telerobotic and inspection. In *Proc. of the 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Yokohama, Japan, pp. 303–309.

McCarthy, M.J. 1990. *An Introduction to Theoretical Kinematics*, MIT.

Michel, D., Marc, R., Jean-Thierry, L., and Gerard, R. 1989. Determination of the attitude of 3D objects from a single perspective view. *IEEE Trans. on PAMI*, 11(12):1265–1278.

Miner, N.E. and Stansfield, S.A. 1994. An interactive virtual reality simulation system for robotic control and operator training. In *Proc. 1994 IEEE International Conference on Intelligent Robotics and Systems*, pp. 1428–1435.

Nakashima, M., Yano, K., Maruyama, Y., and Yakabe, H. 1995. The hot-line robot system "Phase II" and its human-robot interface "MOS". In *Proc. of IEEE 1995 International Conference on Intelligent Robotics and Systems*, pp. 116–123.

Soucy, M. and Laurendeau, D. 1995. A dynamic integration algorithm to model surfaces from multiple range views. *Machine Vision and Its Application*, 8:53–62.

Stenstrom, R. and Connolly, C.I. 1992. Constructing object models from multiple images. *International Journal of Computer Vision*, 9(3):185–212.

Trivedi, M.M. and Chen, C.-X. 1993. Developing telerobotic systems using virtual reality. In *Proc. of the 1993 IEEE/RSJ International*

Conference on Intelligent Robots and Systems, Yokohama, Japan, pp. 352–359.

Wu, C., Wang, D., and Bajcsy, R. 1984. Acquiring 3D spatial data of a real object. *CVGIP*, 28:9.

Wang Jiangang was born in Shanxi, China, on November 5, 1963. He received the B.E. degree in Computer Science in 1985 from Inner Mongolia University, China. In 1988, he received the M.E. degree in Pattern Recognition and Machine Intelligence from Shenyang Institute of Automation, Chinese Academy of Sciences. From 1988 to 1997, he worked at Robotics Laboratory, Shenyang Institute of Automation, Chinese Academy of Sciences where he became an associate professor in 1995. During the academic year 1997 to 1998, he was a research assistant of City University of Hong Kong. He is now working toward the Ph.D. degree in Computer Vision at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His current research interests include computer vision and virtual reality.

Youfu Li received the B.S. and M.S. degrees in Electrical Engineering from Harbin Institute of Technology, China and the Ph.D. degree in Engineering Science from the University of Oxford, UK in 1982, 1986, and 1993 respectively. From 1993 to 1995 he was a postdoctoral research associate in the AI and Robotics Research Group in the Department of Computer Science at the University of Wales, Aberystwyth, UK. He is currently an assistant professor in the Department of Manufacturing Engineering at City University of Hong Kong. His research interests include real-time sensor-based control of robot manipulators, vision guided manipulation, and virtual reality.