PAPER

Semi-Automatic Tool for Aligning a Parameterized CAD Model to Stereo Image Pairs

Chu-Song CHEN[†], **Kuan-Chung HUNG**[†,††], **Yi-Ping HUNG**[†], **Lin-Lin CHEN**[†††], ***and*** **Chiou-Shann FUH**[††], ***Nonmembers***

SUMMARY Fully automatic reconstruction of 3D models from images is well-known to be a difficult problem. For many applications, a limited amount of human assistance is allowed and can greatly reduce the complexity of the 3D reconstruction problem. In this paper, we present an easy-to-use method for aligning a parameterized 3D CAD model to images taken from different views. The shape parameters of the 3D CAD model can be recovered accurately. Our work is composed of two parts. In the first part, we developed an interactive tool which allows the user to associate the features in the CAD model to the features in the 2D images. This interactive tool is designed to achieve efficiency and accuracy. In the second part, 3D information extracted from different stereo views are integrated together by using an optimization technique to obtain accurate shape parameters. Some experimental results have been shown to demonstrate the accuracy and usefulness of the recovered CAD model.

key words: *computer vision, rough CAD model, 3D reconstruction*

1. Introduction

Extraction of 3D information from 2D images has been an important research topic in computer vision. It is well-known that reconstructing dense and accurate 3D depth information is not an easy task. Fortunately, uniformly dense and accurate 3D values are not required for every application. For example, a rough 3D model is sufficient for many applications, such as the following:

1. *Content-based image retrieval*: Usually, in a digital library, we only have the digital images of objects, such as images of antiques and buildings. If rough 3D models for these objects are available, images containing these objects can then be retrieved by query of size, shape, or other 3D attributes of the objects.
2. *Virtual reality (VR)*: 3D information helps a lot in building image-based VR systems. For example, in an object movie system [7], multiple images have to be taken around a 3D object from different views, and these images have to be shown smoothly in rapid succession. The use of 3D information of the object (such as a rough CAD model) not only can allow the interpolated images to be smoother and more realistic, but can also reduce the amount of storage required for the image data.
3. *Augmented reality (AR)*: A fundamental problem in AR is to integrate virtual objects into real images with geometric and motion consistency [2], [6]. The integration will be much easier if rough CAD models of the important objects contained in scene images are available for tracking, localization and rendering.
4. *Image coding*: Two new features of MPEG-4 are the segmentation of images and the composition of 3D graphic models. These features can reduce bit rates and facilitate image syntheses. The segmentation and composition processes can be more efficient if rough 3D models have been reconstructed from images in advance.
5. *Reconstruction of 3D details*: If some good initial estimates of 3D shape can be provided by a rough CAD model, the problem of more detailed 3D reconstruction with stereo vision techniques will become much easier. Because the search range for finding correspondence can be reduced, both the computational speed and stereo correctness can be greatly improved.

Fully automatic reconstruction of 3D models from images is well-known to be a difficult problem. For many applications, a limited amount of human assistance is allowed and can greatly reduce the complexity of the 3D reconstruction problem [8], [9]. In this paper, we describe a semi-automatic tool for reconstructing of rough CAD models from multiple stereo views. This approach can build rough CAD models from images of different views with limited human-assistance. Our approach consists of two stages: the *initial-estimation* stage and the *refinement* stage. The initial-estimation stage can be further divided into two steps: the *human-computer interaction* step and the *stereo reconstruction* step, as shown in Fig. 1. Our human-computer interaction step allows users to interactively associate each edge segment of a rough CAD model (whose shape is fixed but exact parameters are unknown in advance) with that seen in the image. Once the correspondence is established with the human-computer interaction step, the stereo-reconstruction step first generates 3D data of each stereo view, and then register and integrate the

Manuscript received September 28, 1998.
[†]The authors are with Institute of Information Science, Academia Sinica, Taipei, Taiwan.
[††]The authors are with the Department of Computer Science and Information Engineering, National Taiwan University, Taiwan.
[†††]The author is with Graduate Studies in Design, National Taiwan University of Science and Technology, Taipei, Taiwan.

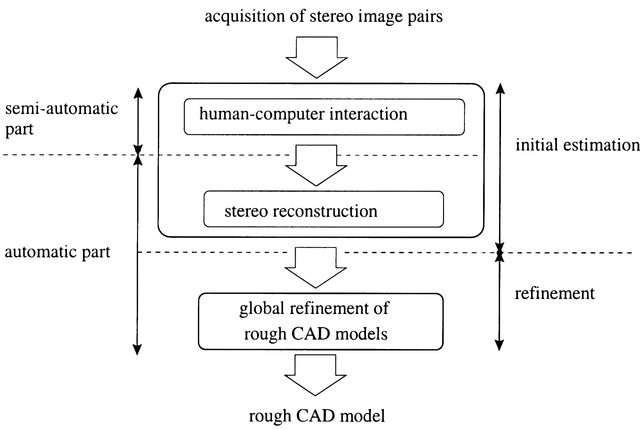


Fig. 1 Block diagram of the semi-automatic tool for building rough CAD models from a set of stereo image pairs.

3D data into a single model. After these two steps, an initial estimate of the parameters of the CAD model is obtained. Finally, in the refinement stage, the model parameters are refined through a nonlinear optimization process. The only semi-automatic part in this approach is the human-computer interaction step at the beginning. All other steps are automatically executed.

This paper is organized as follows. In Sect. 2, we introduce the stage for the generation of initial rough CAD models. In Sect. 3, we describe the refinement of models. Some experimental results are shown in Sect. 4. Finally, Sect. 5 gives some conclusions and discussion.

2. Initial Estimation

2.1 Human-Computer Interaction

The main purpose of our work is to align a rough 3D CAD model to images taken from different views, as shown in Fig. 2. Each image pair is taken by using a calibrated stereo camera set. In this step, we only have to know the topology and *shape constraints* of the 3D CAD model; the scale of each edge of the CAD model is not needed because these will be recovered at later steps.

The model representation in our work is similar to that developed in [8]. A rough CAD model consists of several simple geometric primitives, such as rectangular box, wedge, etc. We call these primitives *shape primitives.* These simple shape primitives are represented in parametric form, with the smallest number of parameters. There are several reasons to represent the 3D model in this way. First, most artifacts or man-made objects are composed of a limited number of regular shapes. Hence, a small number of simple primitives can be combined to achieve a good approximation of the objects. Second, the simpler our model is, the fewer parameters we have to recover. If there are too many parameters to be handled in the system, the optimization process of the refinement stage may take a long time, or even fails sometimes. As shown in Fig. 3, a house can be roughly modeled by a rectangular box together with a wedge. The shape of a rectangular box can be specified by using three parameters (a, b, and c), and the shape of a wedge can be specified by using four parameters (d, e, f, and g).

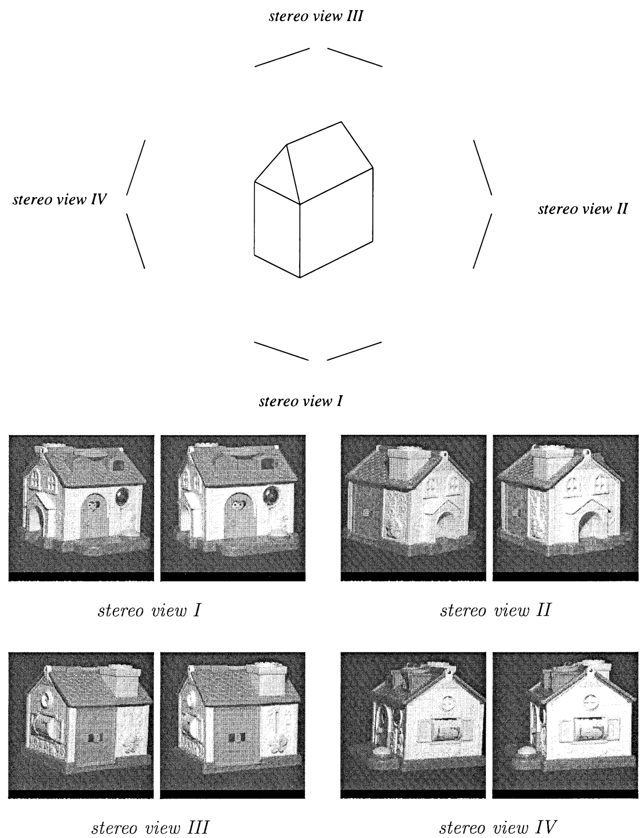


Fig. 2 An example of the multiple stereo views used in the experiments.

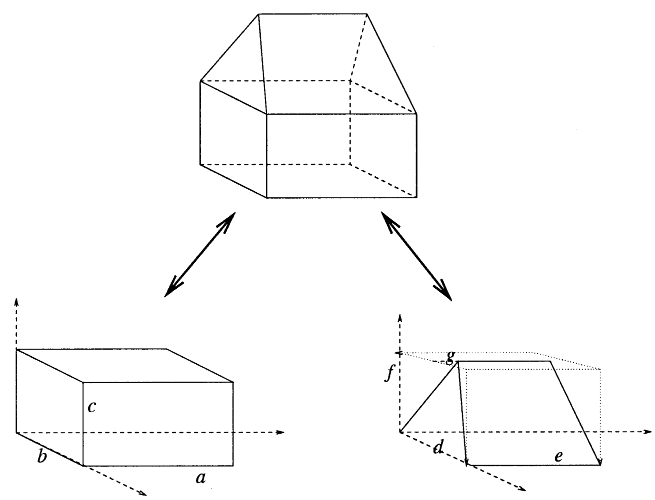


Fig. 3 The parameterized representation of the rough CAD model of a house, which is composed of a rectangular box and a wedge.

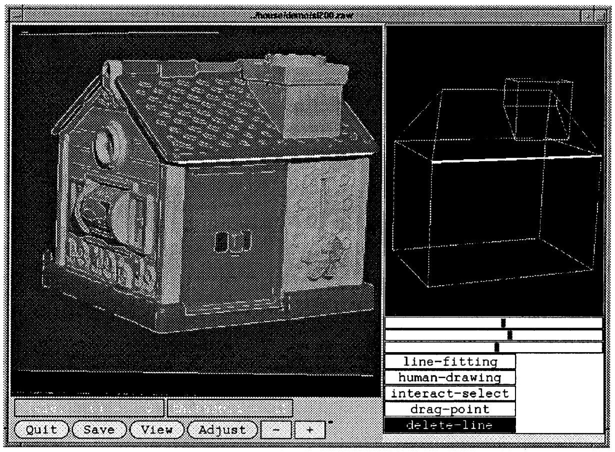


Fig. 4 The graphic user interface for building up the association between the edges in the image and the edges in the CAD model. The left window in the interface is the image window, and the right window is the CAD model window. As illustrated, an edge in the model window is associated with an edge in the image window, and both edges are highlighted.

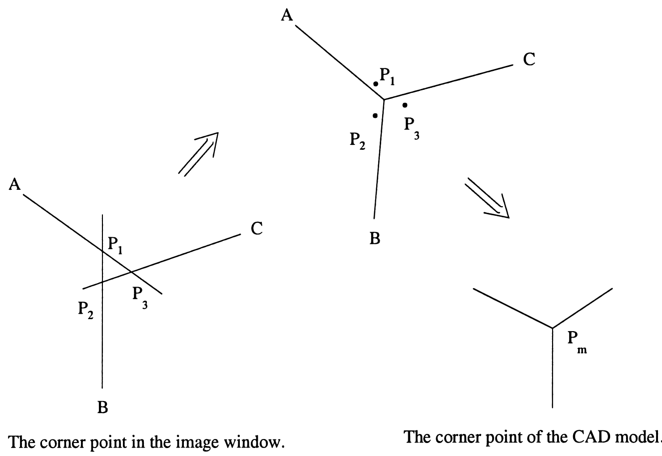


Fig. 5 Three image edges that are supposed to intersect at the same point, as their corresponding edges do in the CAD model, may not intersect exactly at the same point.

Since the precision of the generated object model relies on the precision of the features, we have implemented an easy-to-use interactive tool by which a user can associate edges in the image with those of a CAD model efficiently and accurately. An appearance of the graphical user interface is shown in Fig. 4 in the current implementation, only edge features are considered in the interactive tool. A user can pick an edge and associate it with an edge in the CAD model conveniently. The interface is composed of two windows: *CAD model window* and *image window*. The operations of the two windows are described below:

- The CAD Model Window: After a user builds a CAD model consisting of several shape primitives, the CAD model will be displayed in the CAD model window. Some default values of the shape parameters are assumed only for viewing the model in this step, because the shape parameters (or equivalently, the lengths of all edges) of the CAD model are still unknown. A user can rotate and translate the CAD model to an appropriate viewing position such that the view approximates the image.
- The Image Window: In the image window, an image of the 3D object is displayed. The Canny edge-detection method [4] is used to extract edges in the image, and the detected edges are usually precise, continuous, and thin, as shown in Fig. 4. In addition, a threshold of the gradient values is automatically found, and only strong enough edges (compared to the threshold) will be overlaid. If a user is not satisfied with the threshold, he can interactively adjust the threshold until that he is satisfied with the resulted edges overlaying on the images. Next, he can build up the edge correspondence by first selecting an edge in the CAD model window, and then picking up the correspondent edge in the image window, as shown in Fig. 4. To make the picking more convenient, the interface does not require the user to locate the endpoints of an edge exactly; instead, he can roughly mark a region, and adding or deleting some edge points within this region. Then, least square line fitting will be performed based on the edge pixels collected in this region.

After building up the edge correspondence, the intersection points of the edges have to be computed in our work because the users are not required to pinpoint the endpoints. Since multiple edges may not intersect in a single point, the intersection point is chosen to be the center of all the intersection points of each pair of edges, as shown in Fig. 5. These intersection points are then served as the corner features of the CAD model and are used for 3D reconstruction.

2.2 Stereo Reconstruction

The most difficult problem in 3D reconstruction using stereo vision techniques is the *correspondence problem*. Two typical correspondence problems related to 3D reconstruction are the *stereo correspondence* and the *3D registration*. If the stereo correspondence can be solved, the 3D data of each stereo view can be easily obtained via triangulation because the stereo camera set used in our work has been calibrated in advance. However, the so obtained 3D data is only partial observation, and is represented with respect to a view-dependent coordinate system. Hence, the rotation and translation between adjacent stereo views have to be computed so that these partial 3D data can be integrated into a single data set described with respect to a common coordinate system. To achieve this, we also have to solve the 3D registration problem [3], [5].

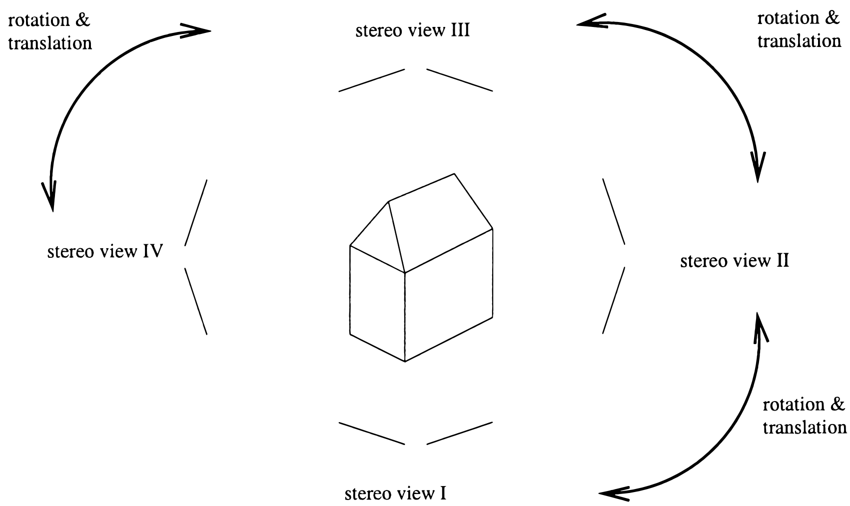


Fig. 6 The registration of adjacent stereo views. The rotation and translation between stereo view I and stereo view IV can be computed by using the rotation and translation between the other three adjacent pairs of stereo views.

First, let us consider the stereo correspondence problem. To the authors' knowledge, no existing method can solve this problem without making mistakes in general cases. Hence, we utilize human-computer interaction to solve this problem semi-automatically. As described in Sect. 2.1, one can build up the correspondence between the edges with images and the edges in the CAD model using our interactive tool. Once the image-model correspondence problem is solved for both the left image and the right image, the stereo correspondence problem is also solved. It is because we can treat the edge segments of the CAD model as *indices* of the image edges of each view corresponding to them. Those image edges corresponding to the same indices should also correspond to each other. For example, in a stereo view, if edge E_l in the left image and edge E_r in the right image correspond to the same model edge E_m, then we know that edge E_l should correspond to edge E_r. Hence, stereo correspondence will be established for E_l and E_r. Similarly, consider two adjacent stereo views. Let E_1 be a 3D edge seen at view 1 (which is reconstructed from the left and right images taken at view 1), and E_2 be a 3D edge seen at view 2. If E_1 and E_2 correspond to the same model edge E_m when solving the stereo correspondence problem, then E_1 and E_2 should also correspond to each other. Hence, 3D registration problem can be solved by establishing up the correspondence between E_1 and E_2.

To save the computation time, 3D reconstruction is performed only for the intersection points of edges (i.e., corners of the CAD model) rather than for all the edge points in the edge segments. In each stereo view, 3D coordinates of the corners of the CAD model can be easily computed by triangulation because stereo correspondence has already been obtained. To find the rotation and translation between adjacent stereo views, the Arun method [1] is used. This method gives a closed form least-squares solutions to the over-constrained 3D-to-3D pose estimation problem. Adjacent views are selected for registration because they have much more overlap. The orientation between non-adjacent views can be computed by composing the rotation and translation matrices between adjacent stereo views, as shown in Fig. 6. Consequently, we have the relative poses between any two stereo views, and the 3D data points obtained with different stereo views can thus be transformed to the same coordinate system.

3. Refinement of Rough CAD Model

A refinement process is required for most image-based interaction systems because human can only make rough associations of the feature positions [9]. In our work, least square line fitting has been used to improve the accuracy of the extracted features, as described in Sect. 2.1. However, the CAD model obtained via the procedures mentioned thus far may not be accurate enough due to the following reasons. First, there will be always some inaccuracy caused by the imprecision of edge detection and by the error in computing the edge intersection points. Second, if the corresponding 3D data points (here, the intersection points) do not align perfectly, we will take a weighted average of the corresponding 3D points as the estimated position of the corners of the CAD model. Hence, the resulted 3D CAD model may be distorted and fails to satisfy the shape constraints of the CAD model. Third, because the rotation and translation are computed for adjacent stereo views, errors may accumulate from one view to another.

An objective function is required to measure the error between our 3D CAD model and the real object. In our work, the error is measured by the sum of errors between the projected edges of the model onto the image and the edges extracted from the image, i.e. $\mathcal{F} = \sum_v \sum_i e_i^v$, where v is the index of image, and i is the index of edges on image v. The computation of e_i^v is illustrated in Fig. 7. In the first case, e_i^v is the area of the quadrangle $abdc$, and in the second case, e_i^v is the sum of the areas of triangle $a'c'e'$ and triangle $b'd'e'$. It is obvious that if the reconstructed CAD model is very accurate, the value of $\mathcal{F}$ should be almost zero. In addition, to avoid the singularity problem caused by the use of Euler angles, we use the quarternion to represent the rotation matrix.

The parameters to be updated in the refinement process can be classified into two types, as shown in Fig. 8. One is the *viewpoint parameters*, i.e., the rotation and translation between adjacent stereo views, which require six parameters (three for translation and three for rotation) for each stereo view. The other is the *shape parameters*, i.e., the parameters used for describing the CAD model (for example, a, b, and c shown in Fig. 8). The number of parameters to be handled in our approach is $6n_v + n_s$, where n_v is the number of stereo views, and n_s is the number of the shape parameters. For example, the case shown in Fig. 8 requires 27(=6×4+3) parameters. To make the refinement process more efficient, the number of the stereo views have to be confined, otherwise, if more stereo views are taken, more parameters are required to be computed in the nonlinear optimization process. However, if the number of stereo views are too small, the overlapping regions of adjacent views may be too small to contain enough common feature parts for performing 3D registration which provides an initial estimate of the rotation and translation between adjacent stereo views. The more complex the model is, the more stereo views the system requires. According to our experience, usually four stereo views are sufficient for the reconstruction of a single object. Given an initial estimate of the viewpoint parameters and the shape parameters, the Newton-Ralphson method in a well-developed numerical tool, the MINPACK package, is used to solve this nonlinear optimization problem.

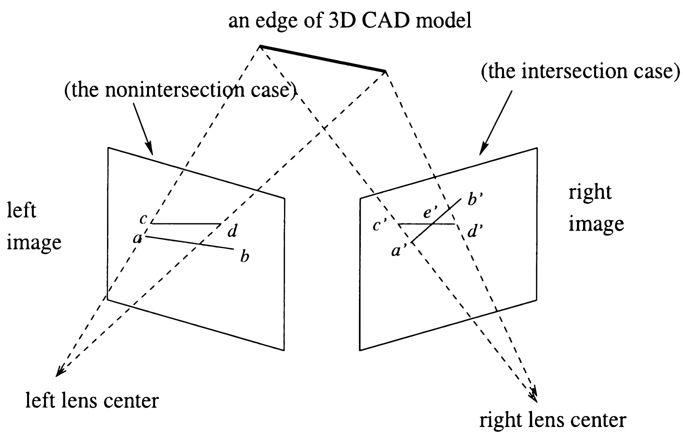


Fig. 7 Definition of e_i^v for two cases (non-intersection and intersection). In the two images, $\overline{ab}$ and $\overline{a'b'}$ are edge segments extracted from the images, and $\overline{cd}$ and $\overline{c'd'}$ are edge segments projected from the edges of 3D CAD model. That the area between $\overline{ab}$ and $\overline{cd}$ is small (the nonintersection case) or the area between $\overline{a'b'}$ and $\overline{c'd'}$ is small (the intersection case) implies that the 3D CAD model is more consistent with the observed images.

4. Experimental Results

To evaluate the quality of the 3D model obtained by applying our algorithms, two experiments are conducted: the first involves the reconstruction of a simple rectangular box; while the second involves the reconstruction of a geometrically more complicated toy house. In

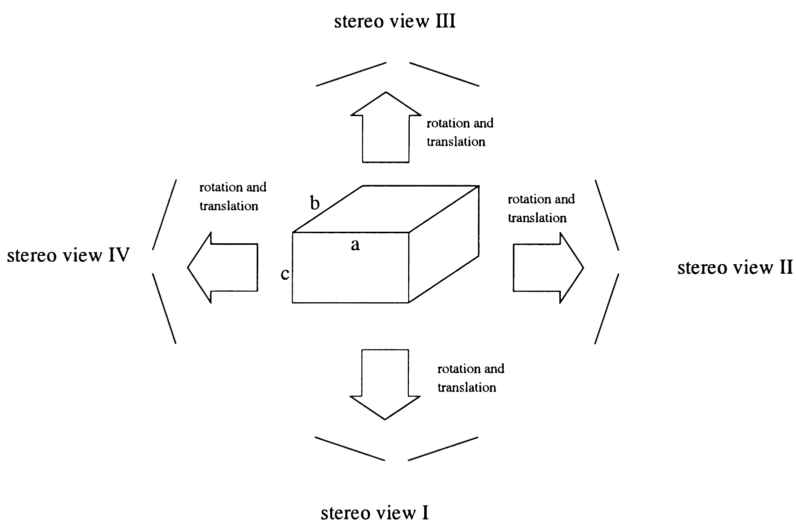
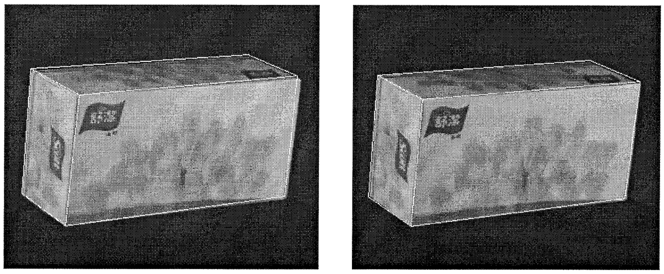
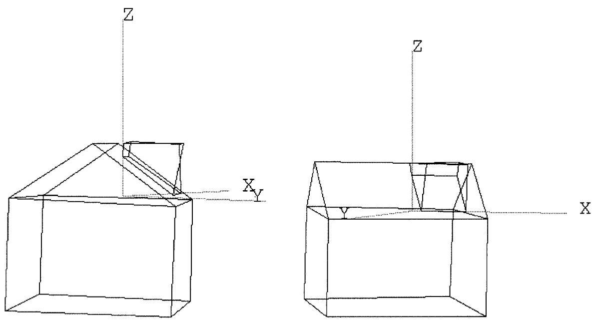


Fig. 8 The type of parameters to be updated in nonlinear optimization: viewpoint parameters (rotation and translation) and shape parameters (e.g., a, b, c).

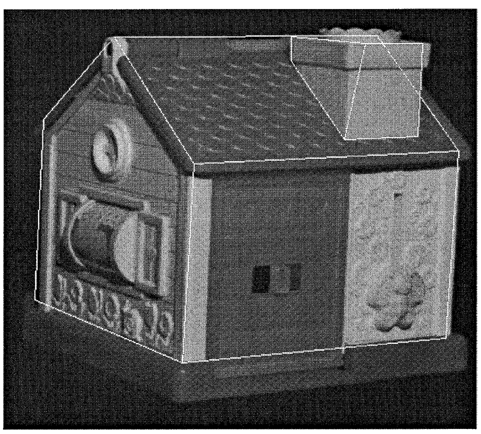


(a) (b)

Fig. 9 (a) The projected lines of the box model obtained from the initial estimation stage. (b) The projected lines of the box model obtained from the refinement stage.



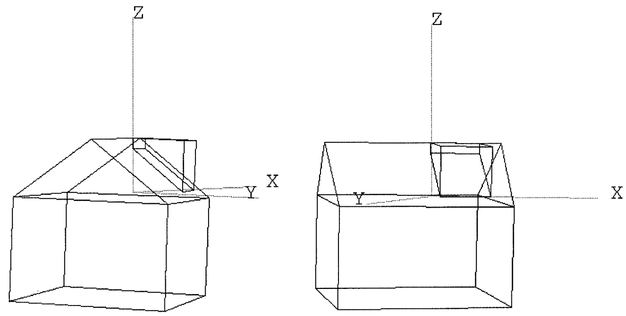
(a) (b)



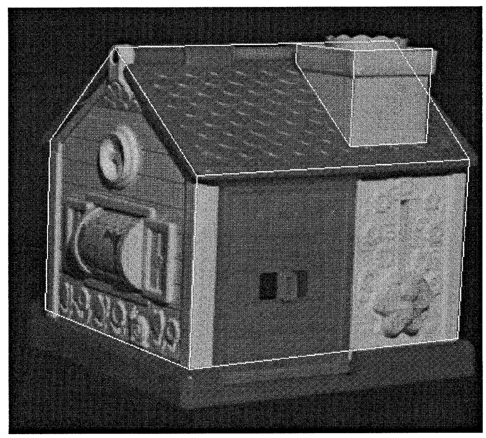
(c)

Fig. 10 (a), (b) Two different views of the wire-frame of the CAD model obtained from the initial estimation stage. (c) The aligned CAD model obtained from the initial estimation stage.

the first experiment, a rectangular box was used for reconstruction. After the association of the edges in the CAD model and the edges seen in each view, partial 3D data can be automatically extracted from different stereo views and integrated into a single model. To see the quality of 3D reconstruction, the CAD model obtained from the initial estimation stage was projected onto an observed image, as shown in Fig. 9 (a). The reconstructed CAD model was aligned with the object seen in the image, but having some distortion at this stage. Then, using the nonlinear optimization process



(a) (b)



(c)

Fig. 11 (a), (b) Two different views of the wire-frame of the CAD model obtained from the refinement stage. (c) The aligned CAD model obtained from the refinement stage.

described in 3, a better 3D CAD model can be obtained (the refinement stage), as shown in Fig. 9 (b). We can see that the shape of the model is undistorted and looks much better than that shown in Fig. 9 (a).

In the second experiment, a toy house was used for reconstruction. The CAD model obtained from the initial estimation stage is shown in Figs. 10 (a)–(c). Because the shape is more complicated, more parameters should be considered in this experiment. It can be seen from Figs. 11 (a)–(c) that the 3D CAD model is refined after applying the global refinement process.

5. Conclusions and Discussion

In this paper, we have developed an easy-to-use tool which can build a rough 3D CAD model from images while involving only limited human-labor. Because it is difficult to solve the 3D reconstruction problem both faultlessly and in an automatic manner, we have developed a semi-automatic tool which is very useful in practical applications, such as those introduced in Sect. 1. Our approach consists of an initial estimation stage and a refinement stage. By using the inherent human capability of perceiving the rough 3D structure from 2D images, a human-computer interface was designed to solve the correspondence problem related to 3D reconstruction. Standard stereo reconstruction and pose estimation methods can then be used to get a good initial estimation of the model. Finally, a nonlinear optimization method is used to globally refine the model. Although there is no new idea in each module used to compose the reconstruction tool described in this paper, we presented a nice example of a working system. It is worth noting that the nonlinear refinement process perform well in our experiments because a good initial guess is made available with our semi-automatic system. The proposed approach only processes line features in this paper (i.e., the CAD models considered in this paper are limited to polyhedra), but the concept presented in this paper can be easily generalized to include the use of curved primitives such as arcs, polynomial curves, splines, and so on.

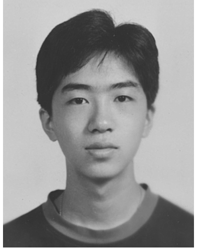
Acknowledgement

This work was supported in part by the National Science Council, Republic of China under Grant NSC 87-2213-E-001-018.

References

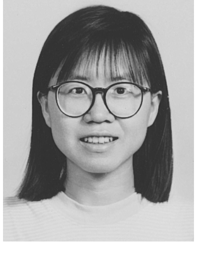
[1] K.S. Arun, T.S. Huang, and S.D. Blostein, "Least-square fitting of two 3-D point set," IEEE Trans. Pattern Anal. & Mach. Intell., vol.9, pp.698–700, 1987.
[2] M.O. Berger and G. Simon, "Robust image composition algorithm for augmented reality," Proc. Asia Conference on Computer Vision, pp.360–367, 1998.
[3] G. Blais and M.D. Levine, "Registering multiview range data to create 3D computer objects," IEEE Trans. Pattern Anal. & Mach. Intell., vol.17, no.8, pp.820–824, 1995.
[4] J. Canny, "A computational approach to edge detection," IEEE Trans. Pattern Anal. & Mach. Intell., vol.8, pp.679–698, 1986.
[5] C.S. Chen, Y.-P. Hung, and J.-B. Cheng, "A fast automatic method for registration of partially-overlapping range images," Proc. International Conference on Computer Vision, ICCV'98, pp.242–248, Bombay, India, Jan. 1998.
[6] C.S. Chen, Y.-P. Hung, S.-W. Shin, et al., "Integrating virtual objects into real images for augmented reality," Proc. ACM International Symposium on Virtual Reality Software and Technology, VRST '98, pp.1–8, Nov. 1998.
[7] S.E. Chen, "QuickTime VR—An image-based approach to virtual environment navigation," SIGGRAPH'95, Los Angeles, CA, 1995.
[8] P.E. Debevec, C. Taylor, and J. Malik, "Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach," SIGGRAPH'96, pp.11–20, New Orleans, Louisiana, 1996.
[9] Y.P. Hung, C.-S. Chen, K.-C. Hung, et al., "Multi-pass hierarchical stereo matching for generation of digital terrain models from aerial images," Machine Vision and Applications, vol.11, no.1, pp.280–291, 1998.

Chu-Song Chen is an assistant research fellow in the Institute of Informaton Science, Academia Sinica, Taiwan. His research interests are in computer vision and virtual reality. He received his B.S. from National Chiao-Tung University, Taiwan, in 1989, and his Ph.D. from National Taiwan University in 1996.



Kuan-Chung Hung received his B.S. in 1994 and M.S. in 1996, both from the Department of Computer Science and Information Engineering, Taiwan. He is now working for the EyeArt Corporation, Taiwan.

Yi-Ping Hung received his B.S. in electrical engineering from National Taiwan University in 1982. He received a Ph.D. from the Division of Engineering at Brown University, in 1990. He then joined the Institute of Information Science, Academia Sinica, Taiwan, and became a research fellow in 1997. He served as the deputy director of the Institute of Information Science from 1996 to 1997, and received the Outstanding Young Investigator Award of Academia Sinica in 1997. Dr. Hung has published more than 70 technical papers in the fields of computer vision, pattern recognition, image processing, and robotics. In addition to the above topics, his current research interests also include visual surveillance, virtual reality, and visual communication.



Lin-Lin Chen is a professor of industrial and commercial design with National Taiwan University of Science and Technology. Her research interests are in computer-aided design, computational geometry, and web-based applications. She received her B.S. from National Cheng Kung University, Taiwan, in 1984 and her Ph.D. from University of Michigan, U.S.A., in 1992.

Chiou-Shann Fuh received the BS degree in computer science and information engineering from National Taiwan University, Taipei, Taiwan, in 1983, the MS degree in computer science from the Pennsylvania State University, University Park, PA, in 1987, and the Ph.D. degree in computer science from Harvard University, Cambridge, MA, in 1992. He was with AT&T Bell Laboratories and engaged in performance monitoring of switching networks from 1992 to 1993. Since 1993, he has been an associate professor in the Computer Science and Information Engineering Department at National Taiwan University, Taipei, Taiwan. His current research interests include digital image processing, computer vision, pattern recognition, and mathematical morphology.