A Next-Best-View System for Autonomous 3-D Object Reconstruction

Joseph E. Banta, Laurana M. Wong, Christophe Dumont, and Mongi A. Abidi
Department of Electrical Engineering
The University of Tennessee, Knoxville

Abstract

The focus of this paper is to design and implement a system capable of automatically reconstructing a prototype three-dimensional model from a minimum number of range images of an object. Given an ideal 3-D object model, the system iteratively renders range and intensity images of the model from a specified position, assimilates the range information into a prototype model, and determines the sensor pose (position and orientation) from which an optimal amount of previously unrecorded information may be acquired. Reconstruction is terminated when the model meets a given threshold of accuracy. Such a system has applications in the context of robot navigation, manufacturing, or hazardous materials handling.

The system has been tested successfully on several synthetic data models, and each set of results was found to be reasonably consistent with an intuitive human search. The number of views necessary to reconstruct an adequate 3-D prototype depends on the complexity of the object or scene and the initial data collected. The prototype models which the system recovers compare well with the ideal models.

I. Introduction

When a system encounters an unknown object or environment, the system may be required to reconstruct a prototype or model of the object. Consider the case of an autonomous robot manipulating objects in its environment. The robot must have knowledge of the volume and position of an object in order to pick it up. If no *a priori* information is available about the object, the robot must be capable of acquiring data from which a working model of the object may be constructed.

The objective of this work is to build a system that constructs the best possible 3-D model of a static object. This involves acquiring range information from a number of sensor locations around the object, aligning the range data with the 3-D world coordinate system, and then updating the model with the new information. The reconstruction system must then inspect the updated 3-D model and determine, using a pre-specified set of decision rules or heuristics, the next range sensor position and orientation which will reveal the greatest amount of unknown 3-D information. Finding sensor locations that obtain the greatest amount of previously unknown information regarding an object or scene is referred to as the next-best-view

(NBV) problem (see [1], [2], [3], [4]). The ultimate product of a system which solves the NBV problem is a list of a minimal number of sensor positions which allow for complete reconstruction of an object or scene.

Our system renders range and intensity images of an object from each next view position and assimilates the range data into the existing prototype model. Three different approaches are used to determine the next best sensor position from the current state of the model. We use a coarse algorithm during the initial stages of reconstruction and switch to finer and finer approaches as reconstruction progresses. New view positions are calculated until the system decides to terminate reconstruction. The final output of the system is a volumetric model and surface rendering of the target object. We have tested our reconstruction system on two object models.

In this work, we assume that the size of the object is roughly known. Therefore, we can use a *view sphere* to limit the number of possible sensor positions and simplify calculations (see Figure 1). The radius of the view sphere is large enough to encompass the entire object of interest, and the center of the view sphere is aligned with the approximate center of the object. In this work, we align the view sphere with the coordinate axes of our computer models. For each new view acquired, the sensor is always positioned toward the center of the view sphere.

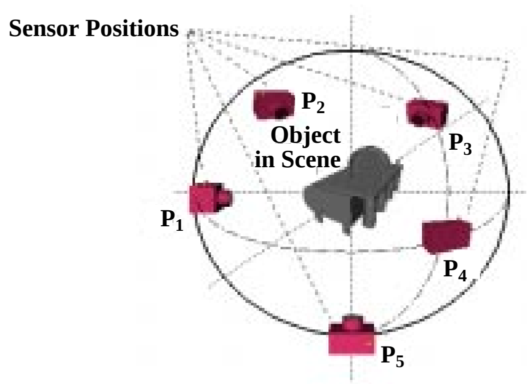

Fig. 1. *The reconstruction procedure involves assimilating range data from various positions into a world model.*

The remainder of this paper is arranged as follows: Section II provides background information on range imaging, occupancy grids, and data registration. Section III addresses the NBV problem, including a review of related research and implemented methods. Section IV desribes the integration of the complete NBV system, and Section V contains results for two 3-D models. Finally, Section VI summarizes our research findings and states conclusions about the system results. Statements are also made regarding future work and potential improvements to the current system.

II. Background

The following sections give background information on three subjects that are fundamental to our next-best-view system. Sections II-A and II-B briefly review range imaging and occupancy grids. Section II-C discusses the methodology used to incorporate range data into a 3-D representation.

A. Range Imaging

Range is defined as the distance from a point in a scene to a point in a known reference coordinate system. Raw range data, in the form of a range image, only provides us with $2\frac{1}{2}$-dimensional information; the range image describes only the surfaces ranged from a specific position and orientation. In order to extract true 3-D scene information, we require knowledge of the camera pose. We define the camera pose as the position and orientation of the camera in world coordinates. Knowing the pose of the camera, we can compute a rigid transformation that will allow us to register newly acquired range data with our world model. (See [5] for more information on range imaging.)

B. The Occupancy Grid

The occupancy grid is a 3-D array in which each element represents a specific volume in the discrete tessellation of the workspace. The grid may be viewed as a simple extension of the 2-D bitmap. Just as each pixel in a 2-D bitmap can take on one of two binary values, each voxel in a 3-D bitmap (occupancy grid) can take on a value of 0 (unoccupied) or 1 (occupied). ([6], [7], and [8] contain more information on 3-D modeling.)

C. Range Data Registration

To register range data, we must first transform the range data from $2\frac{1}{2}$-D to 3-D by calculating the transformation from one coordinate space to another. We can easily shift a greyscale range image into three dimensions by separating the grey levels into binary bit planes and occupying the voxel corresponding to each pixel and appropriate bit plane. We then apply a transformation that maps a ranged point to its corresponding Cartesian coordinates.

To restrict view positions to points outside of the model and to ensure model visibility, the sensor is always placed on the view sphere and oriented toward the approximate center of the object. Because the range images are all of the same resolution and acquired at the same radius from the approximated object center, neither the translation nor the scale factor component of the transformation are necessary.

Once we have calculated the transformation matrix, the range data may be registered. We create a *world occupancy grid* from the surface data in each new range image and the hidden information which lies occluded from the camera's view (the occluded data). An illustration of our approach is shown in Figure 2. ([9], [10], [11], [12], and [13] provide more information on range data registration.)

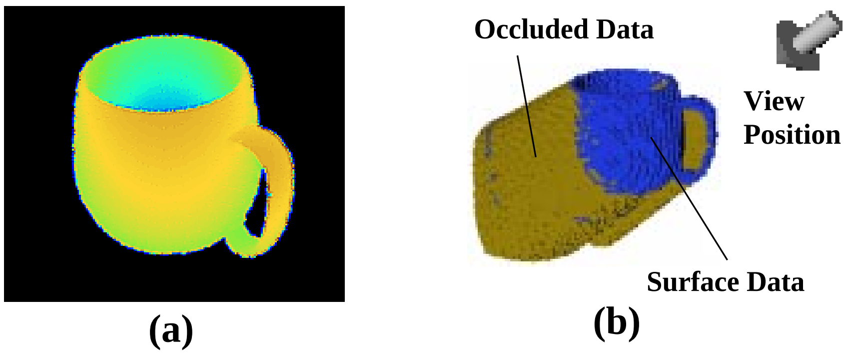

Fig. 2. *(a) A range image, and (b) a model reconstructed from the surface data contained in (a) and the occluded data inferred from (a).*

III. The Next Best View

The following sections address the next-best-view problem. Section III-A reviews papers that are most relevant to our topic. Section III-B discusses the three approaches that we developed for determining the next best sensor position.

A. Review of Related Work

In this section, we present a literature review which examines the topics in sensor planning that are most relevant to the NBV problem. Many of the principles to be described in the reviewed documentation were adopted in this paper.

Abidi [14] focuses on the use of a vision sensor to perform the volumetric modeling of an unknown object in an entirely autonomous fashion. An entropy-based objective functional is used to model the object or scene in areas of occluding contours. A local maximum of this function is found to determine the best next view position.

Some of the first work in automated range sensor positioning was done by Connolly in 1985 [1]. In this research, Connolly develops two algorithms for determining the next best view using partial octree models. In the first algorithm, the next view position is the point at which the greatest surface area of unseen nodes is visible. The second algorithm uses information about the node faces in the octree which are common to both unseen and empty nodes.

Maver and Bajcsy [2] describe an automated occlusion-guided view determining system for scene reconstruction. The system is given a priori knowledge of the environment and sensor geometry and returns the next view position from which a complete range image of the surface visible to the camera may be obtained. The system then computes new scanning planes for further 3-D data acquisition based on the discontinuities (occlusions) in the most recent range image. Maver et al. [3] propose a NBV system using the max-min principle as a heuristic. The system selects from all possible viewing directions the one which maximizes the amount of new information.

Pito and Bajcsy [4] present a system that automatically acquires a surface model of an arbitrary part. The concept of positional space is introduced as a basis for view representation. Two types of information are recovered from range data: the visible surface and the void surface. The criterion for determining the next view position is set by the void surface area visible from a point. The 'visible and void' idea is similar to the concept of 'surface and occluded' data that we introduced in Section II-C.

For more information on sensor planning, Taranabis et al. [15] have compiled a survey of sensor planning in computer vision. Work is reviewed for feature detectability using generate-and-test, synthesis, and expert systems. The authors describe the work in sensor planning as the initial efforts of researchers to address a problem with many degrees of freedom. It is noted that future work is required for incorporating more realism into the object models and relaxing some of the constraints made on developed systems.

B. Methods for Determining the Next Best View

In the following sections, three approaches for determining the NBV are discussed. The first two sections introduce simplistic approaches to sensor placement which produce adequate results and find justification in human intuition but have inherent limitations. The last section introduces a more localized approach to sensor placement. The algorithms are listed in a coarse to fine order; the coarsest algorithm is presented first, the second algorithm takes a finer approach to finding the next best view, and the last algorithm is the finest approach. The integration of all three methods is described in Section IV-D.

It is important to note that the following methods were developed with a concern for the simplicity of implementation and the quality of the returned result. Also, each algorithm calculates only one point on which to focus the sensor. The next view is simply the 3-D vector which intersects both the point of focus and the center of the view sphere.

B.1 Edge-Based Sensor Placement

This approach for positioning the range sensor reveals large areas of occlusion. This method is useful in the initial stages of reconstruction when only one range image has been acquired. A similar approach was developed by Maver and Bajcsy [2]. In this method, the range image is examined for evidence of occlusions which might conceal surface information. We can detect occlusions by measuring the level of intensity of the edges in the range image. We assume that the higher intensities in the first derivative edge map conceal greater quantities of unknown information. Therefore, the intuitive choice for the next view direction is given by the brightest intensity of the edge map and the center of the view sphere (see Figure 3).

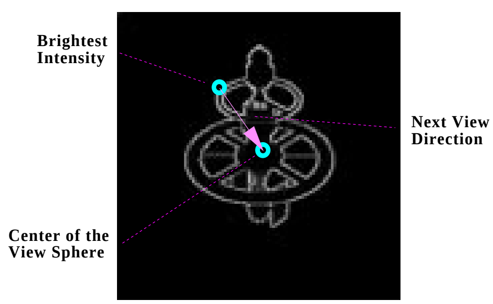

Fig. 3. *The next view direction intersects the brightest intensity of the edge map and is oriented toward the center of the view sphere.*

B.2 Targeting the Occluded Surface Centroid

In this approach, we examine the reconstructed voxel model. Since we would like to view as much of the occluded surface area as possible, we position the sensor toward the centroid of an occluded surface. The new view points toward the view sphere center and is oriented on the voxel face closest to the centroid of the occluded region. This approach performs well when confronted with one continuous occluded surface as in Figure 4(a). The centroid of that surface is surrounded by information which would prove useful in reconstructing a 3-D model. However, in the case of Figure 4(b), where the occluded surface is divided into several different areas, the centroid approach might position the sensor at a seemingly ambiguous location. This leads us to conclude that the centroid approach would perform well during the initial view acquisitions, but its performance would degrade as reconstruction progresses. We will take this issue into consideration when integrating the methods into our NBV system.

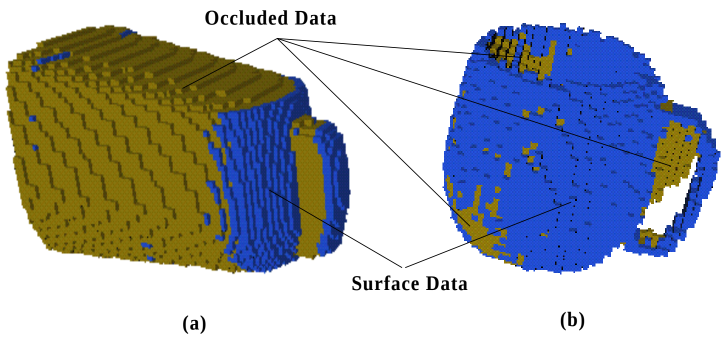

Fig. 4. *(a) A model in which the exposed occluded surface is fairly continuous, and (b) a model in which the exposed occluded surface is composed of several disjoint patches.*

B.3 Clustering the Occluded Surface Data

Every approach for sensor placement that we have proposed up to this point performs well during initial view acquisitions, but none of them localizes the search for occluded information. This localization is necessary during later view acquisitions when the occluded surface is broken up into several disjoint surface patches as in Figure 4(b). The obvious conclusion is that we should develop a sensor positioning technique to use during later view acquisitions which isolates some of the more significant surface patches.

We have chosen a clustering technique to divide the occluded patches into an unspecified number of areas. The clustering approach we have chosen to develop is a variation of the *minimum distance classifier* [16]. Initially, all occluded patches belong to the same cluster. The mean or *prototype* of each cluster is calculated. Each voxel face is made a member of the cluster with the nearest prototype. If the face farthest from its cluster prototype is half as far from its prototype as the mean distance from another prototype, it is made the member of a new cluster. The process repeats until the voxel face farthest from its cluster prototype is less than half as far from its prototype as the mean distance from another prototype.

This clustering approach tends to find clusters which are closely packed and, thus, is well suited for separating disjoint surface patches. The next sensor position points toward the view sphere center and is oriented on the voxel face closest to the mean of the largest cluster found by the clustering algorithm.

IV. System Integration

The simplicity of the three NBV algorithms just described requires us to integrate them into a larger system. For this integrated system to perform autonomously, there are several issues that must be discussed. Section IV-A considers the situation in which the next-best-view system produces a view position from which the area of interest is not visible to the sensor. Section IV-B discusses the situation in which the system places the sensor unnecessarily close to a previous sensing location. Section IV-C defines the criteria which determines when the system has adequately reconstructed a model. In Section IV-D , we assimilate the developed approaches into an autonomous NBV system. Section IV-E presents the software written to implement the complete NBV system.

A. Providing the Sensor with an Unobstructed View of a Feature

For those sensor placement approaches which determine view positions based on the state of the current model (the centroid and clustering methods), we must assure that the area of interest is actually visible from the specified viewpoint. This issue arises particularly when we attempt to reconstruct an object with self-occluding concavities. An example of such an object is illustrated in Figure 5.

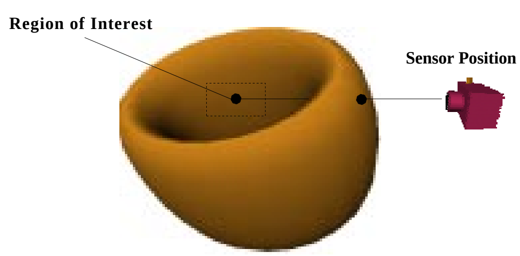

Fig. 5. *An object containing a self-occluding concavity.*

If a ray from the view direction to the target area intersects any occupied voxels in the reconstructed object model, a view of the region will be occluded. The solution to this problem is to slightly perturb the angle of the view direction.

If we transform the Cartesian view direction vector into spherical coordinates, we have a vector (r,θ,ϕ) comprised of a radius and two angles. Holding θ constant, we first iterate ϕ through all angles within 90^o

of the original view direction, in increments of one degree. The algorithm checks the status of all voxels on a line with the calculated direction radiated from the target point. If an occupied voxel is encountered, the next angle is checked. If no unoccluded directions are found, θ is incremented by one degree, and ϕ is iterated through all angles within 90^o. In this fashion, we iterate ϕ through all angles within 90^o. If no unoccluded direction is found at this point, the algorithm fails. This will not happen often because the region of interest is exposed will probably be visible from some direction.

B. Separating Closely-Spaced Views

The NBV system may give a sensor position which is unnecessarily close to a previous sensor position. Figure 6 illustrates such a situation.

Fig. 6. *A situation in which the NBV system has determined a sensor position which is unnecessarily close to a previous sensor position.*

The intuitive solution to this problem is to push the new view position away from the previous position. Therefore, we want to set a minimum angle between the new view vector and each previous view vector. The situation may exist, however, that a user does not care if sensor positions are relatively close to one another. For this reason, we will make the view separation function optional, and it may be engaged by the user at any time.

C. Criteria for Reconstruction Termination

For a system to automatically reconstruct a 3-D model, it must be capable of self-termination. We would like for the reconstructed surface area to be substantially larger than the occluded surface area. This will be true if the model has been sufficiently reconstructed. However, we should continue reconstruction if a relatively large amount of new data is being acquired with each view. Therefore, we should also take into consideration the surface area of the model after the acquisition of views.

Thus, the system terminates reconstruction if the ratio of surface faces to occluded faces is large *and* the change in surface face count is small (less than 10%). Reconstruction also terminates if the change in occluded face count is small (less than 10%).

D. A Combination of Next-Best-View Methods

While the criteria for algorithm termination was established in Section IV-C, we still have not completely automated the system. In order for the NBV system to be autonomous, it must be capable of selecting the *best* method for determining the next view position at the current stage of reconstruction. In the previous sections, several such methods were introduced and discussed, including the strengths and weaknesses of each method. We use a coarse approach during the initial stages of reconstruction and switch to finer and finer approaches as reconstruction progresses.

The position from which the first image is acquired is arbitrary; It will either be set to a default position or it may be specified. So that the second view vector will always be orthogonal to the first, the second view position is determined using the edge-based technique of Section III-B.1. For the intermediate third and fourth view positions, we use the centroid approach of Section III-B.2. When reconstruction is in this stage, the occluded surfaces will probably be fairly continuous, so the overall centroid of the occluded surface should yield a good deal of previously unknown information.

Theoretically, the best method to use during the latter stages of reconstruction is the clustering approach of Section III-B.3. However, it is possible that the approach, when applied repeatedly, might orient the sensor on uninformative regions of the object surface. An autonomous system should safeguard against this possibility. One way to do this is to monitor the rate at which reconstruction is proceeding. If reconstruction is proceeding well, we can assume that the clustering approach is performing as expected and that it may be applied to the acquisition of another view. If, however, reconstruction seems to have stalled, we can assume that there is a problem and that another approach is needed. In this case, we orient the sensor as far as possible from all previous positions. One way to accomplish such a task is to find a point on the restricted view sphere which maximizes the sum of the distances from the previous views. We find the new view direction by taking the average of the previous view directions and multiplying by -1 to position the new view on the opposite side of the view sphere. In this way, we can assure that reconstruction will always continue normally. A flowchart of the procedure followed by the system is illustrated in Figure 7.

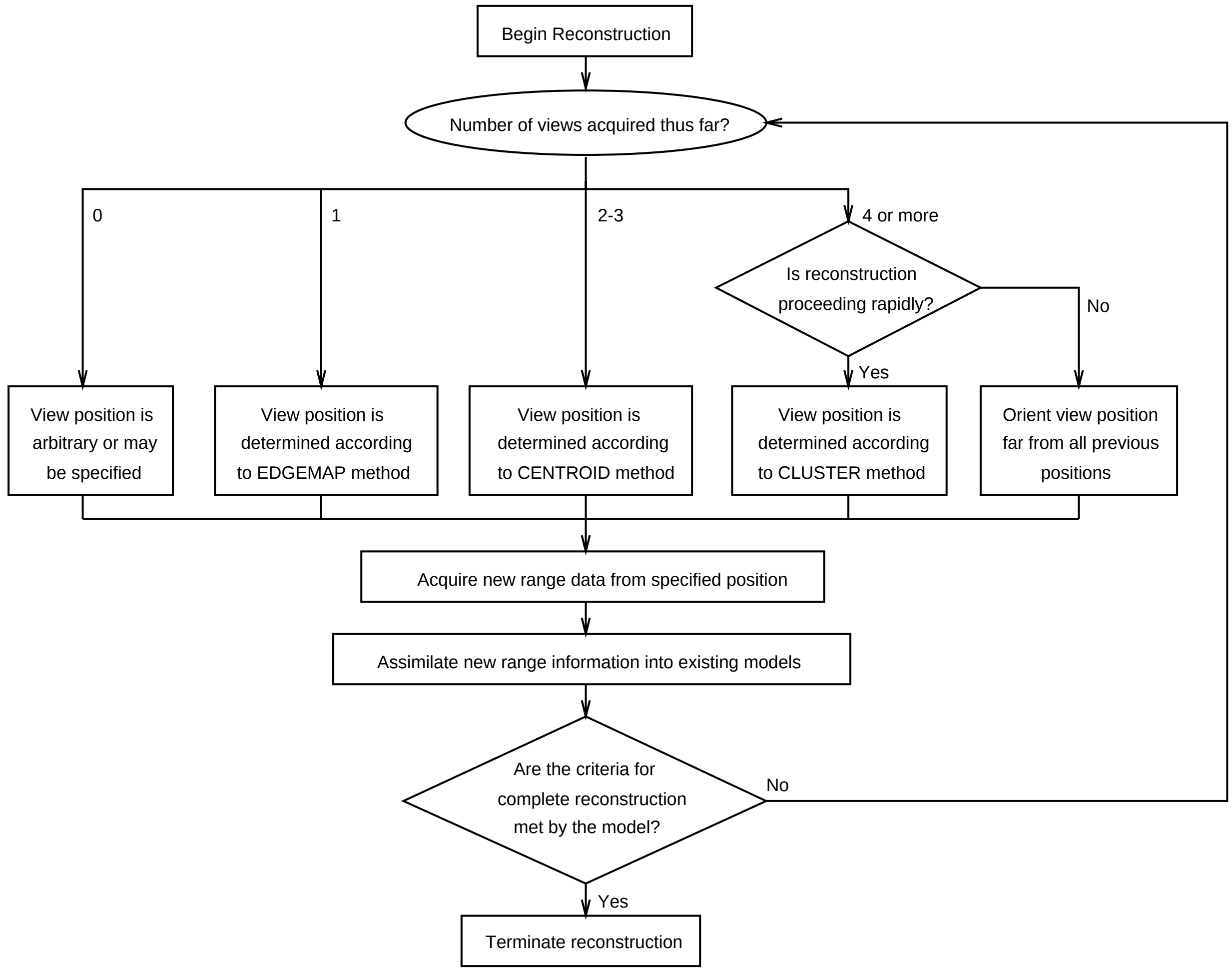

Fig. 7. *A flowchart of the next-best-view system procedure.*

E. Software Implementation

We have written a C^{++} program to implement the next-best-view system. On the command line, one can specify the name of the file containing the ideal model and the name of the file containing the reconstructed model. The dimensionalities of the acquired images and the reconstructed model may be designated, as well as the encoding type of the output image files. The application is capable of determining the next view position, acquiring the range/intensity image pair from the specified view position, and reconstructing the model from the most recently acquired range information.

The user may also control the methodology used in reconstructing the prototype model. The coordinates of the next view position may be manually set, or one may specify the method to be used. The user may tell the system how many views to acquire using the specified method. The option also exists to let the system reconstruct the model as completely as possible. If this option is selected, the model is reconstructed to the threshold of accuracy as discussed in Section IV-C, and the method used to determine the next sensor pose is determined by the rules discussed in Section IV-D. One can also have the system separate closely spaced views, as discussed in Section IV-B.

An option also exists for the launch of a graphical user interface. This interface was written in Java externally to the ‘nextview’ C^{++} code and is simply a front end for the actual system. Screen shots of the graphical interface and its four interdependent panels are illustrated in Figure 8.

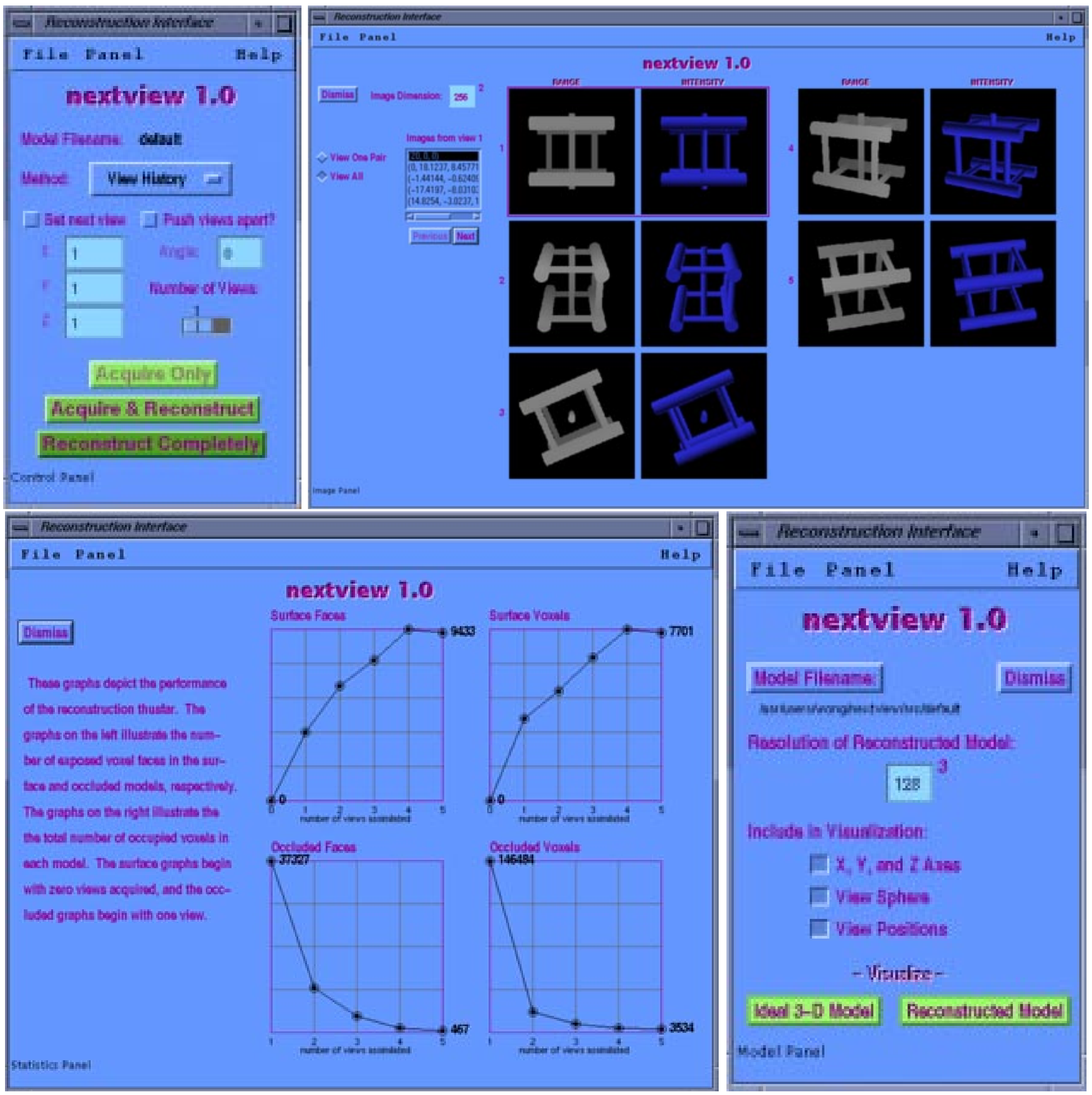

Fig. 8. *Clockwise from top left: the Control Panel, the Image Panel, the Model Panel, and the Statistics Panel of the graphical interface.*

The graphical interface presents all the same options to the user as the command line. In addition, the user is able to view (on the Image Panel) all range and intensity images acquired thus far, review the statistical behavior of the reconstruction by examining the graphs on the Statistics Panel, or spawn an external viewer (from the Model Panel) to visualize either the ideal 3-D model or the reconstructed model. Online help is also available which contains programmer information, the application's purpose, and documentation regarding the use of the 'nextview' application and its interface.

V. Experimental Results

In this section, we present results from the NBV system. The standard resolution of the acquired images is 512x512 , and the resolution of the reconstructed models is 256x256x256. At this level of model resolution, the limited complexity of the models allows for adequate tessellation of the work space. The time required for reconstruction is typically between two and three minutes for the determination, acquisition, and assimilation of each view on a 195 MHz Silicon Graphics R10000 Indigo2 Impact/TRAM workstation. The acquisition and assimilation of the views accounts for most of the reconstruction time. The time to calculate the next best view is minimal. We present results from the reconstruction of two models, the pipe model and the hand model.

A. The Pipe Model

The pipe model, shown in Figure 9, is an array of cylinders arranged in a grid pattern such that the model contains many self-occluding contours. Thus, this model should adequately test the method for providing an unoccluded view of a target vicinity on an object.

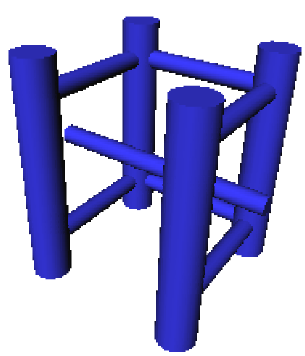

Fig. 9. *The ideal pipe model.*

The first range and intensity images of the pipe model, shown in Figure 10, are acquired from an arbitrary position. A voxel rendering of the recovered pipe model after the assimilation of the first range data set is shown in Figure 11. In this illustration, the dark shaded surface is the recovered surface data, and the brighter shaded surface is the occluded data. We continue reconstruction from this model.

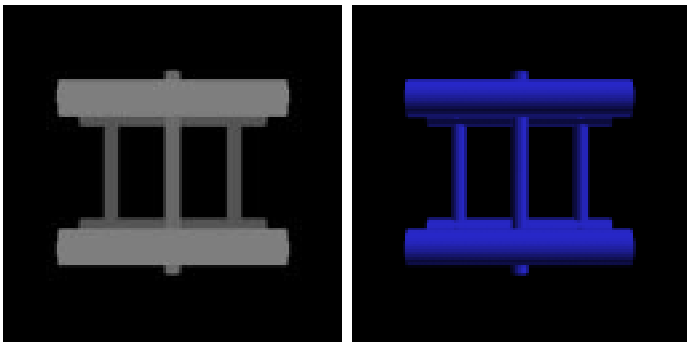

Fig. 10. *Range and intensity images acquired from the first view position for the pipe model.*

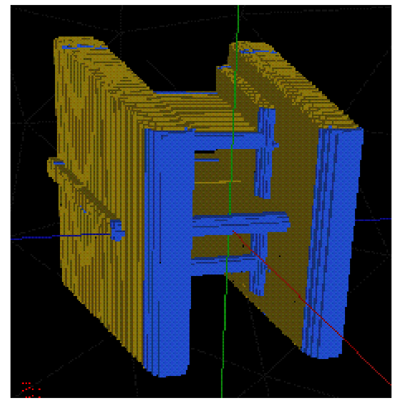

Fig. 11. *A voxel rendering of the recovered pipe model after assimilation of the first range data set.*

The second view position is determined by the edge-based technique of Section III-B.1. The range and intensity images acquired from the second view position are shown in Figure 12, and a voxel rendering of the recovered model after the assimilation of the second range data set is shown in Figure 13. Notice that the amount of occluded information has been greatly reduced between the first and second views. This would probably be true of any two view positions separated by a reasonable angle but because these two positions are necessarily orthogonal, the amount of new information recovered is significant.

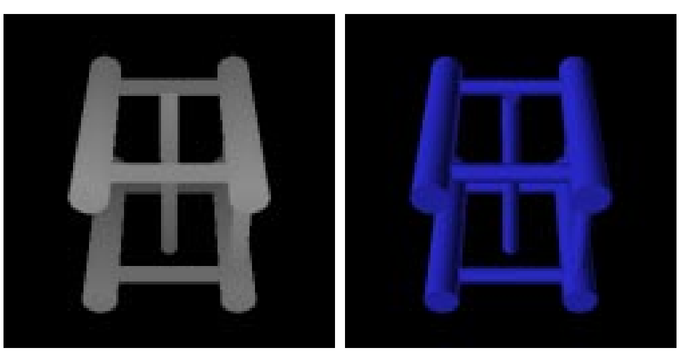

Fig. 12. *Range and intensity images acquired from the second view position for the pipe model.*

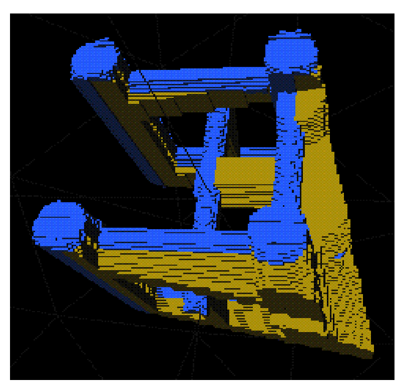

Fig. 13. *A voxel rendering of the recovered pipe model after assimilation of the second range data set.*

The third view position is determined using the centroid method described in Section III-B.2. Figure 14 shows the range and intensity images acquired from the third view position for the pipe model. Note that the third view position is nearly orthogonal to both the first and second view positions. For this reason, the amount of new information recovered from the third range data set is also fairly significant.

As we can see from the voxel rendering of the recovered model after three views (Figure 15), reconstruction is nearing completion. The occluded data has been broken into several disjoint surface patches at an early stage, which would lead us to select the clustering method to determine the next view position. However, since the NBV system automates the reconstruction process, the method chosen is again the centroid approach. For this example, the model is being reconstructed more rapidly than originally expected. We, however, would not expect the same speed of reconstruction for a more complex model.

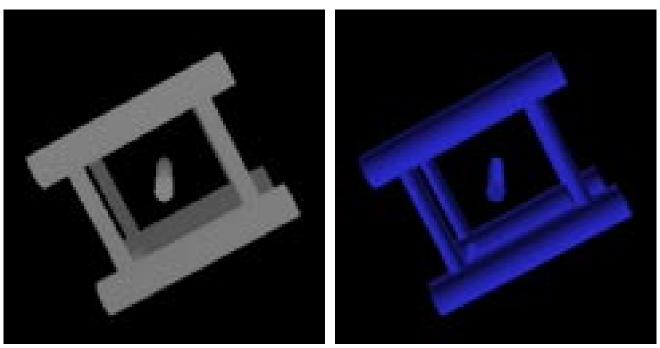

Fig. 14. *Range and intensity images acquired from the third view position for the pipe model.*

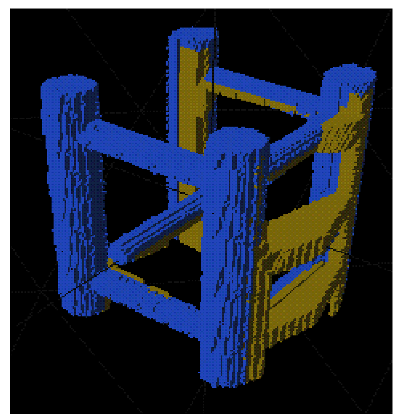

Fig. 15. *A voxel rendering of the recovered pipe model after assimilation of the third range data set.*

Luckily, the centroid approach orients the fourth sensor position on the two large clusters of occluded data visible in Figure 15. Figure 16 shows the range and intensity images acquired from the specified position, and a voxel rendering of the recovered pipe model after the assimilation of the fourth range data set is illustrated in Figure 17. The model after four view acquisitions seems fairly complete. Only a small number of scattered occlusions remain, and these will probably be recovered with the next view acquisition.

Because reconstruction has been proceeding rather rapidly, the NBV system determines the location of the fifth view position using the clustering approach as discussed in Section III-B.3. The results of the clustering algorithm are circled in Figure 17. The next view position is selected by the system to eliminate

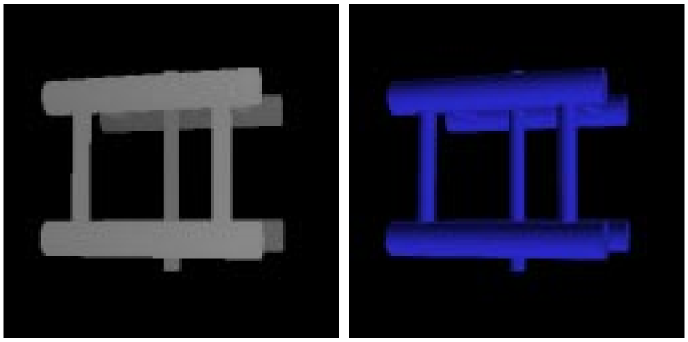

Fig. 16. *Range and intensity images acquired from the fourth view position for the pipe model.*

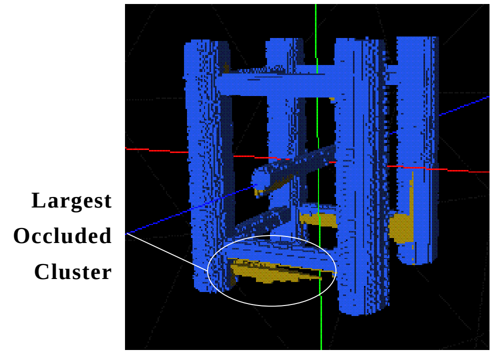

Fig. 17. *A voxel rendering of the recovered pipe model after assimilation of the fourth range data set (with the largest occluded cluster circled).*

this occluded cluster, and the images acquired from this position are shown in Figure 18. Figure 19 shows a voxel rendering of the reconstructed model after the assimilation of this range data set. The largest occluded cluster from Figure 17 was eliminated, and it appears that any remaining occlusions in the model are sparse. After the assimilation of the fifth range data set, the NBV system decides that the model is adequately reconstructed.

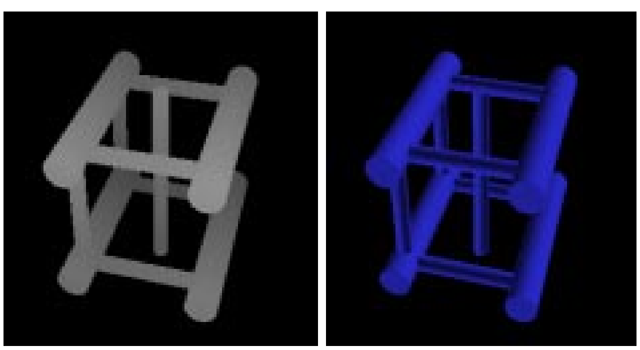

Fig. 18. *Range and intensity images acquired from the fifth view position for the pipe model.*

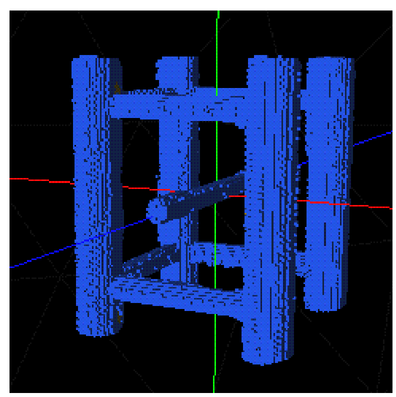

Fig. 19. *A voxel rendering of the recovered pipe model after assimilation of the fifth range data set.*

Now that the NBV system has reconstructed a prototype of the pipe model, we may examine the collective results. Figure 20 shows the ideal pipe model and each of the acquired view positions. In Figure 21, we see a surface rendering of the model compared with the ideal pipe model. One can see that the reconstructed model visually resembles the ideal model. The surface of the reconstructed model is not as smooth as the surface of the ideal model since our volume data is binary.

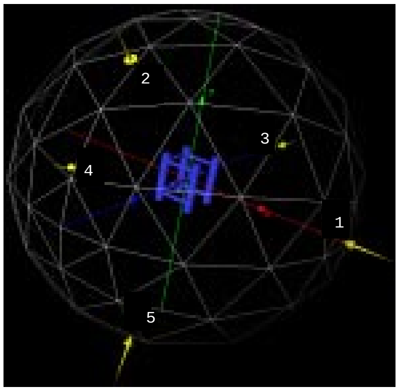

Fig. 20. *The ideal pipe model shown with the acquired view positions (represented as arrows).*

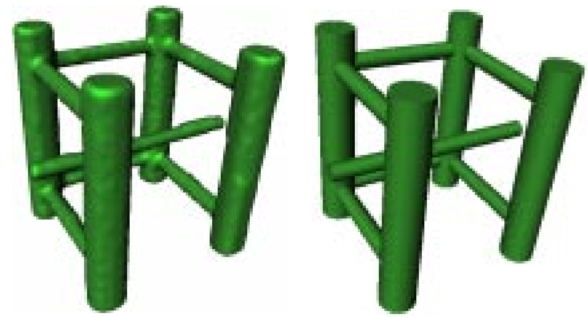

Fig. 21. *A surface rendering of the reconstructed pipe model (left) compared with the ideal pipe model (right).*

B. The Hand Model

The ideal hand model is shown in Figure 22. This model is a robotic manipulator with three finger-like appendages and a semispherical base. Located in the center of the robot arm are three cylindrical concavities which should provide a challenge to the reconstruction system. The near-negligible thickness of each finger with respect to the overall size of the model may introduce an additional obstacle for the system to overcome.

After the acquisition and assimilation of four range views, the criteria for termination are met by the reconstructed model. A voxel rendering of the reconstructed hand model is illustrated in Figure 23.

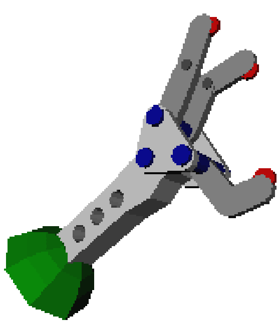

Fig. 22. *The ideal hand model.*

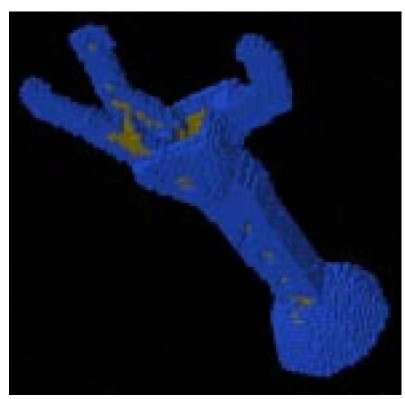

Fig. 23. *A voxel rendering of the hand model after four range image assimilations.*

While the criteria for sufficient reconstruction were met, it appears that the cylindrical concavities in the center of the arm were not correctly reconstructed, and a fair amount of occluded information remains to be assimilated near the base of the fingers. For this reason, four more range images are acquired using the view history approach. The set of eight range and intensity images acquired for the hand model are shown in Figure 24. Figure 25 shows the view positions relative to the ideal hand model.

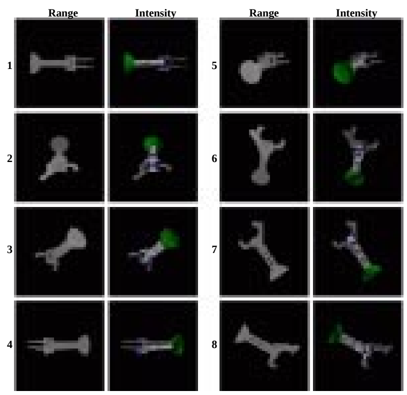

Fig. 24. *The eight range and intensity images acquired for the hand model.*

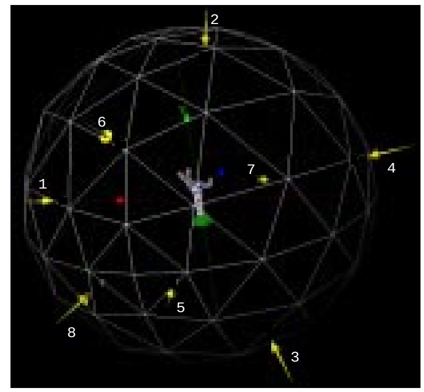

Fig. 25. *The ideal hand model with the eight view positions shown.*

Figure 26 shows four graphs which depict the behavior of the NBV system after eight views. The graphs on the left show the number of surface and occluded voxel faces. The graphs on the right show the surface and occluded voxels. The surface graphs (top) show that the exposed number of surfaces does increase in size with the assimilation of each range data set until reconstruction termination. The occlusion graphs (bottom) illustrate the dwindling size of occlusions with each view.

While the number of exposed occluded voxel faces converges toward zero, the number of occluded voxels will always converge to some finite positive number. These are the voxels *inside* of the model. Also, the number of occluded voxels decreases to the internal voxel count of the model, while the number of surface voxels peaks and then gradually trails off. This is explained by those voxels in the surface model which are essentially 'false positives': those voxels which appear from one range image to be on the surface of the object, but are later verified from another image as not being occupied volumes at all.

If we examine a voxel rendering of the discrete surface of the reconstructed model after eight range assimilations (Figure 27), we see that many of the occlusions which were visually identified in Figure 23 have been assimilated into the reconstructed prototype. Figure 28 shows an isosurface rendering of the reconstructed hand model compared with the ideal hand model. Once more, the reconstructed model visually compares well with the ideal model, and, thus, extending reconstruction for this model was appropriate.

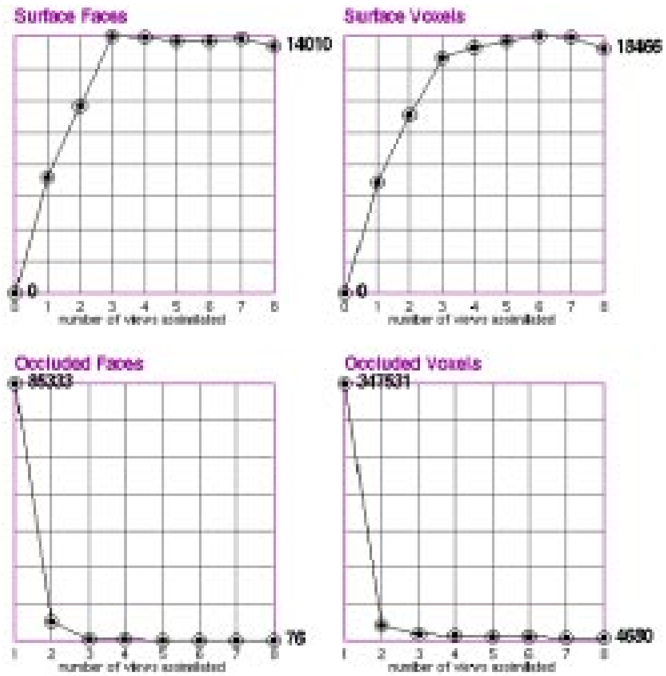

Fig. 26. *Four graphs depicting the size of each model as reconstruction progresses.*

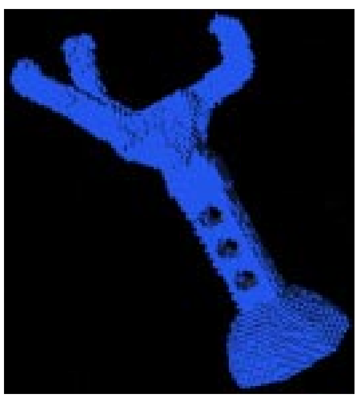

Fig. 27. *A voxel rendering of the reconstructed hand model after the assimilation of eight range views.*

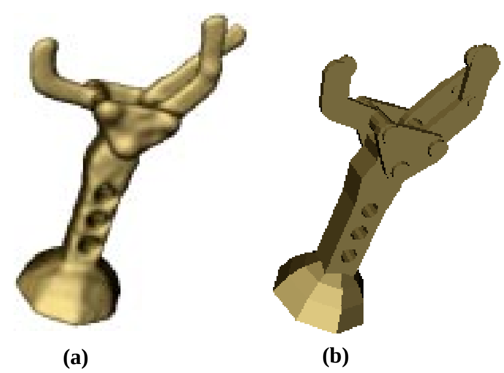

Fig. 28. *An isosurface rendering of the reconstructed hand model (a) compared with the ideal hand model (b).*

VI. Summary and Conclusions

After reviewing the results, the overall performance of the next-best-view system has surpassed our original goals. The system is able to acquire range and intensity images from a specified position with respect to a given model, assimilate the range information into a 3-D bitmap model of a given size, and determine from the current status of the model the position from which an optimal amount of new range information can be acquired from the model. The system emulates the intuitive inspection and reconstruction capabilities of a human being. The reconstructed models, depending on the object complexity and the resolution of the model, are consistently accurate and robust reproductions of the original ideal model of the object.

The digital nature of the voxel data and the resolutions of the models themselves prevent the reconstructed models from replicating the ideal models exactly. A model of very large resolution would present the most accurate reproduction of any model, but we are limited by practicality and available facilities. Considering the scope of the problem, the NBV system performed well. The application should be suitable for implementation in a practical situation such as in the modeling of an actual 3-D object.

The next step is to implement the system in the real world; Plans exist for the system to be applied to the automated modeling of a facility contaminated with radioactivity. This is the situation in which a disposal team has been assigned to dismantle a facility contaminated with hazardous waste. The team must acquire information regarding the layout of the facility and the locations of any potential hazards. If no a priori information is available, the team must employ a system capable of acquiring information from which a working model of the facility may be constructed. With such a model, tele-operated disposal machinery can be maneuvered within.

The input range images would be acquired by a Perceptron P5000 laser range scanner to be mounted on a Schilling Titan II manipulator and supervised by a dedicated Sun SPARC 5 workstation. All elements of this proposed system exist, are functional, and have only to be compiled into one working system.

A definite possibility for the NBV system in the future will be to work with probabilistic models rather than binary models. While the current software works with models which have one bit of memory allocated for each voxel in the 3-D model, the system could easily be altered to work with models which have an entire byte of memory allocated for each voxel in the prototype model. A probabilistic model would improve the accuracy of isosurface visualization. Another possibility is to work with octrees and use higher resolution in areas of greater physical contrast.

Optimally, the system should limit which positions may be considered for the next best view. An actual system would probably not be able to view an object from any angle. Thus, the NBV system should be modified to register range data from any location from the object.

In conclusion, we should note the many separate components generated specifically for this project:

1. The C^{++} class library, complete with operators and operator functions, written for handling and manipulating discrete three-dimensional models.

2. The application which reads an IRIS Inventor format file and renders range and intensity images of the described object from a given position.
3. The registration function which assimilates the existing prototype model and the most recently acquired range image.
4. The NBV system which presents three separate methods for determining the NBV position from the current state of the model.
5. The graphical user interface with which the user can call the NBV system, view the images acquired at each iteration, review statistics pertaining to reconstruction, or examine the ideal or reconstructed model.
6. The application which reads in a reconstructed model file and outputs an IRIS Inventor format voxel rendering.
7. The IRIS Explorer module map which reads in a reconstructed model file and outputs an IRIS Inventor format surface rendering.

References

[1] C. I. Connolly, "The determination of next best views," *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 432–435, 1985.

[2] J. Maver and R. Bajcsy, "Occlusions as a guide for planning the next view," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 417–433, May 1993.

[3] J. Maver, A. Leonardis, and F. Solina, "Planning the next view using the max-min principle," in *Computer Analysis of Images and Patterns*, pp. 543–547, Springer-Verlag, September 1993.

[4] R. Pito and R. Bajcsy, "A solution to the next best view problem for automated cad model acquisition of free-form objects using range cameras," *Proceedings of the SPIE*, vol. 2596, pp. 418–429, October 1995.

[5] R. A. Jarvis, "A perspective for range finding techniques for computer vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, pp. 122–139, April 1993.

[6] Y. Chen and G. Medioni, "Surface description of complex objects from multiple range images," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 153–158, 1994.

[7] C. H. Chien, Y. B. Sim, and J. K. Aggarwal, "Generation of volume/surface octree from range data," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 254–260, 1988.

[8] O. D. Faugeras and M. Hebert, "The representation, recognition, and positioning of 3-d shapes from range data," *Techniques for 3-D Machine Perception*, pp. 13–51, 1986.

[9] E. Bittar, S. Lavallée, and R. Szeliski, "A method for registering overlapping range images of arbitrarily shaped surfaces for 3-d object reconstruction," *Proceedings of the SPIE*, vol. 2059, pp. 384–395, 1993.

[10] H. Gagnon, M. Soucy, R. Bergevin, and D. Laurendeau, "Registration of multiple range views for automatic 3-d model building," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 581–586, June 1994.

[11] K. Higuchi, M. Herbert, and K. Ikeuchi, "Building 3-d models from unregistered range images," *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2248–2253, May 1994.

[12] M. Rutishauser, M. Stricker, and M. Trobina, "Merging range images of arbitrarily shaped objects," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 573–580, June 1994.

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

[14] B. A. Abidi, "Automatic sensor placement," *Proceedings of the SPIE*, vol. 2588, pp. 387–398, October 1995.

[15] K. A. Taranabis, P. K. Allen, and R. Y. Tsai, "A survey of sensor planning in computer vision," *IEEE Transactions on Robotics and Automation*, vol. 11, pp. 86–104, February 1995.

[16] R. C. Gonzalez and R. E. Woods, *Digital image processing.* Addison-Wesley Publishing Company, 1992.