A Robotic System for 3-D Model Acquisition from Multiple Range Images

Michael K. Reed and Peter K. Allen
Computer Science Department
Columbia University, New York, NY 10027

Abstract

This paper describes a robotic system that builds a 3-D CAD model of an object incrementally from multiple range images. It motivates the generation of a solid model at each stage of the modeling process, allowing the use of well-defined geometric algorithms to perform the merging and integration task. The data from each imaging operation is represented by a mesh, which is then extruded in the viewing direction to form a solid model. These solids are merged as they are acquired into a composite model of the object. We describe an algorithm that builds a solid model from a mesh surface and present experimental results of reconstructing a complex object. In addition, we discuss an approach to completely automating the model acquisition process by integration with previous sensor planning results.

1.0 Introduction

Three dimensional computer models have become the cornerstone for an increasing number of applications in robotics and other fields. In robotics, these models have been used for assembly planning, mobile robot navigation, grasp planning, reverse engineering, and accurate simulation of robotic workcells, each of which requires a method of acquiring models of real-world objects or scenes. The problem of acquiring these models has been termed *modeling from observation* [7]. This problem is difficult for several reasons. One is the large scope of the geometry and topology of general 3-D objects, which includes both concave and convex polygonal surfaces, curved surfaces, and features such as holes. Another problem is the size of the data sets that are acquired, which may require extremely large computational and memory resources. This is complicated by the difficulty of determining the number and location of views to reduce the number of sensing operations and the data size.

This paper describes a robotic system that incrementally builds solid 3-D models from multiple range images. It motivates the generation of a topologically correct 3-D solid model at each stage of the modeling process. This allows the use of well-defined geometric algorithms to perform the merging and integration task, which is one of the more difficult phases of the modeling process. A major goal of this system is to retain object fidelity while reducing the number of scans through the use of a sensor planning process.

Early research on this task included work based on intensity images, but the focus has recently shifted to utilizing range images due to the increased availability of accurate rangefinders. The REFAB system allows a user to specify approximate locations of machining features on a range image of a part; the system then produces a best fit to the data using the previously-identified features and domain-specific knowledge as constraints [13]. The IVIS system uses an octree to represent the seen and unseen parts of each of a set of range images and uses set-theoretic operators to merge the octrees into a final model [12]. Methods that use a mesh surface to model and integrate each of a set of range images [16] [11] or to model a single, complete point sampling [6] have also proven useful in this task. Very recently, an octree-based modeler has been used as input to an isosurface extraction routine, resulting in a closed mesh model [5].

While the majority of work in this area assumes that the images cover enough of the object to build an acceptable model, there has been progress towards adding a planning component to avoid user specification of the imaging operations, reduce the number of views, and improve the model's fidelity. This planning component has the potential to simplify the model acquisition process. This has implications for applications such as 3-D FAX, where objects are automatically acquired at one site, transmitted over a network, and rebuilt at a remote site using Rapid Prototyping technology.

The method for model acquisition described in this paper uses a mesh to model the sensed surface of an object, and then sweeps the mesh in the imaging direction to generate a solid representation. In this regard it may be thought of as an integration of both the mesh-based methods described above and previous work that performs edge detection and projection from intensity images [4][9]. The models created by this method differ from previous mesh-based methods in two important regards. First, the model created from each imaging operation is that of a solid which permits it to be used by CAD/CAM or planning packages which expect a closed model, and also allows rough models that may be acceptable to some tasks to be created in as few as two orthogonal sensing operations. This is in contrast to mesh surface patches that do not form a closed model until the entire scanning process is completed, thereby precluding any planning system that relies on a closed model. Thus, our method is an incremental one that allows new information to be easily integrated as it is acquired into a composite model. Second, each model created by our method includes information about the volume of occlusion, which is not present in systems that only model the object's surface. The *volume of occlusion*, which is the space occluded from the sensor during each imaging operation, has previously been used to help guide the planning process [3] [8] [10]. The occlusion volume is a key component of many sensor planning methods because it allows the system to reason about what *has not* been seen, but it has not yet been integrated into mesh-based methods. This paper motivates its use and construction for mesh-based models.

Our system iterates through 4 stages until a satisfactory model is built. First a range image of the object is acquired from one sensing position, or *viewpoint*. A mesh is then constructed to model the range image, resulting in a mesh surface. This surface is then swept or extruded to form a solid model from a particular viewpoint that is integrated into the current model. The system then determines the next viewpoint and returns to the first step. Currently, the viewpoints are known *a priori*, but we present a method by which the partially built models may be used to effectively determine the next viewpoint during the acquisition process. The remainder of this paper discusses each of these phases in turn.

2.0 A robotic system for acquiring range images

Our goal is to create a fully automated model acquisition system. As such, a robotic system for acquiring a range image of the object being modeled has been built. This system is comprised of a Servo-Robot laser rangefinder attached to an IBM 7575 SCARA robot (see Figure 1), with the object to be imaged being placed on a motorized rotation stage. This stage allows sensing from different viewpoints by rotating the part instead of moving the robot and rangefinder, and increases the accuracy of the modeling process. After the rangefinder acquires one stripe of M points parallel to the world XY plane, the robot steps the rangefinder in the world -Z direction, and the process repeats N times until a complete NxM image is made. Point data from the rangefinder is transformed into the coordinate frame of the initial rotation stage orientation by using the known transformations between the rangefinder, robot, turntable, and current turntable rotation.

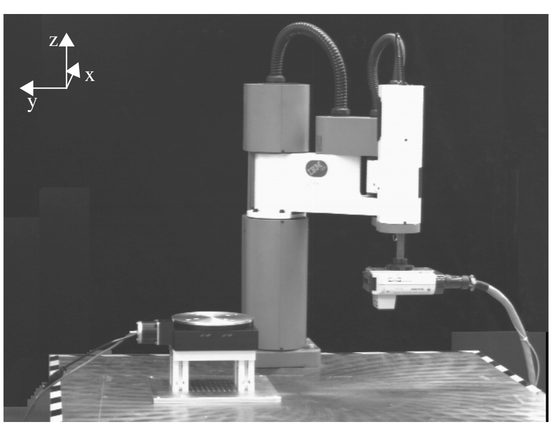

FIGURE 1. Experimental setup showing robot with attached laser rangefinder (to right) and nearby rotation stage (to left). World coordinates are annotated in upper left.

3.0 Building the mesh

A mesh is a piecewise linear surface composed of elements that meet along their edges, which in turn meet at vertices. Meshes are frequently chosen to represent a sampled surface due to their efficiency, their representational flexibility, and the simplicity of mesh algorithms. They find particular application in range imaging where objects are highly oversampled during the sensing process. Mesh surfaces built from these range images may then be efficiently processed to reduce their size, fit with more complex surface types, or registered to each other. However, since the mesh determined by a single-view range image is in essence a surface model, it does not contain information that permits spatial addressability (the ability to classify points as inside, on, or outside the model) which is necessary for many tasks, as do solid models. Although a mesh that completely covers an object may be used to determine a solid model, in most incremental modeling techniques the mesh can not be closed until the end of the scanning process. This precludes the use of a planning method or any other procedure that requires a solid model.

A solution to this problem is to build a solid model from each scanning operation that incorporates both the information about the models surfaces (the sensed data) and occlusion information in the form of the occlusion volume (the unsensed data). When building the mesh that will be used to represent a surface from a range image, it is necessary to determine both the points from the range image that will be included and what the mesh connectivity will be. Each data point in the range image becomes a vertex in the mesh, and a simple 6-connectivity is used for the mesh edges. In this regard our work differs from other mesh-based methods such as mesh zippering [16] and other similar re-meshing techniques [11] which retain only elements that lie directly on an imaged surface by removing elements that have an edge whose length exceeds some threshold (see Figure 2.). Our

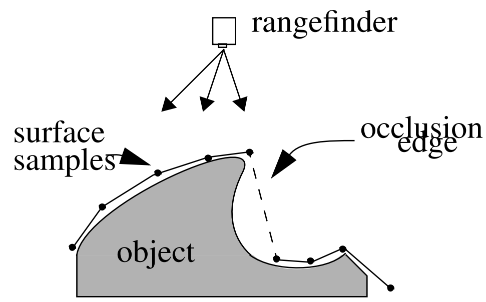

FIGURE 2. Example of edges between sampled vertices on a surface.

system retains these elements, since they denote parts of the surface that are occluded from the sensor and need further imaging, and therefore are useful in the planning process. We do, however, remove elements that lie in the background of the image via a threshold for reasons of efficiency.

As an example of this process, consider the hypothetical object shown at the top of Figure 3. A synthetic range image is sampled from the CAD model using the shown sensing direction. The surface (middle row, Figure 3)is typical of mesh-based methods; no occlusion volume is represented, and hence it is unsuitable for the technique presented here. In contrast, the mesh at the bottom of Figure 3 represents both the imaged surface of the object and the occluded regions.

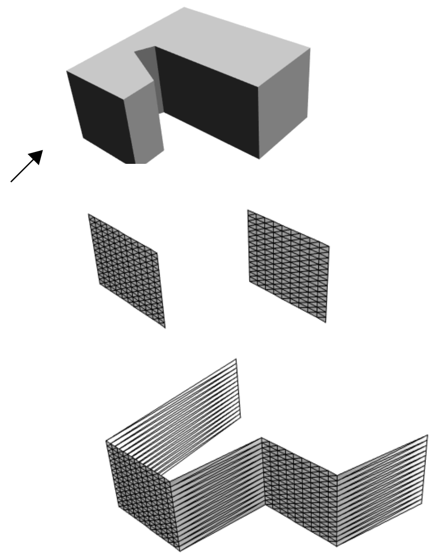

FIGURE 3. Top: rendering of CAD model of a typical 2-1/2 D part, shown with a sensing direction. Middle: surface mesh from synthetic range data of above part. This mesh does not include any elements from the volume of occlusion. Bottom: surface mesh generated from synthetic range data, including volume of occlusion.

4.0 Sweeping the mesh into a solid

Once a mesh of the input image has been constructed, the surface is swept to extrude a solid model of both the imaged object surfaces and the occluded volume. The individual surfaces that make up the mesh are swept in the viewing direction and are then integrated by performing a set union operation (see Figure 4).

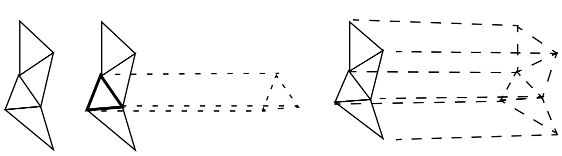

FIGURE 4. Example of a mesh sweep operation. (left to right) Mesh surface, mesh surface with one element swept, and mesh surface with all elements swept and unioned. The sensing direction is from the left.

Each triangular mesh element is swept orthographically along the vector of the rangefinder's sensing axis until it comes in contact with a far bounding plane, resulting in the 5-sided solid of a triangular prism, as shown in the example above. This may be done either by using projective geometry or by using the extrusion facility available in most CAD packages. Each solid may then be combined with the others by a regularized set union operation. Because the sweeping direction is the identical for each element, each prism constructed from a non-boundary mesh element (i.e. an element that shares each edge with another element, as opposed to those that are not surrounded by 3 other elements) will share its lateral faces and edges with those of a neighbor, allowing a simpler and more efficient algorithm than a full-featured union operation to be used.

As an example, consider again the hypothetical part shown in Figure 3. Sweeping its mesh results in the solid shown in Figure 5.

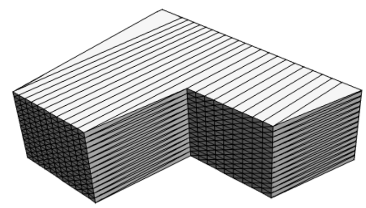

FIGURE 5. Solid formed by sweeping the mesh shown at the bottom of Figure 3 **in the sensing direction.**

The polyhedral solid swept out consists of 3 sets of surfaces: a mesh-like surface from the acquired range data, a number of lateral faces equal to the number of vertices on the boundary of the mesh derived from the sweeping operation, and a planar bounding surface that caps one end. It no longer adheres to the form of the input mesh in that it includes non-triangular surface elements at both the lateral faces and capping surface. For use in planning, it is important to be able to differentiate between these surfaces during later model analysis. One way to do this is to attach tags to each surface in the model based on which of the above sets the surface belongs to. All surface elements in the model that were present in the mesh before sweeping and that are composed of edges shorter than a threshold distance should be tagged as "imaged surface", in accordance with the description in [3]. These elements describe surfaces of the object that were imaged properly and do not need to be imaged again. All the remaining surfaces should be tagged as "occluded volume" so that they may be used to drive a later planning process. It should be noted that this tagging procedure must be done to a model from a single viewpoint: after models have been merged large faces often get split into smaller ones during the merging process, and will not be differentiable by their edge lengths alone. After the tagging process the solid may be merged with models from other viewpoints, or it may first be used as input to a mesh optimization routine to reduce the number of elements.

5.0 Merging single-view models

Each successive sensing operation will result in new information that must be merged with the current model being built. In prior research merging of mesh-based models has been done using clipping and re-triangulation methods. These methods were necessary because the meshes constructed from each viewpoint are not closed, and because of this they are specialized to operate on non-manifold surfaces of approximately continuous vertex density. An advantage of our method is that, because we generate a solid from each viewpoint, we may use a merging method based on set intersection included in the solid modeling capabilities of CAD systems. Many modern CAD systems include highly robust algorithms for set operations on solids, which is of critical importance in this application for the following reasons: the high density of the range images (and therefore the small size of many of the mesh elements), the many long and thin lateral surfaces, and most importantly the fact that many of these models will have overlapping surfaces that are extremely close to each other. Finally, because the single-view and merged models should be 2-manifold, it is necessary to use set operations that are able to handle regularized intersection.

The merging process itself starts by initializing the sensed "composite" model to be the entire bounded space of our modeling system. The information determined by a newly acquired model from a single viewpoint is incorporated into the composite model by performing a regularized set intersection operation between the two. In should be noted that the intersection operation must be able to correctly propagate the "imaged surface" and "occluded volume" tags from surfaces in the models through to the resulting model. Because surfaces may be split or deleted during the intersection operation, this may be a non-trivial problem. In this work we have used Spatial Technology's ACIS geometric modeler, which is able to handle both merging at the required level of detail and the tag propagation.

6.0 Experimental results

We present here an example that demonstrates the capabilities this system by building a composite model from distinct views. We will build a CAD model of the object shown in Figure 6, which is a strut-like part. This part has smooth and polygonal surfaces, and includes holes that are not imagable in our current sensor configuration. Four 110x128 range images, take from equi-angular turntable rotations, are shown in Figure 7. Following the method described here, a mesh is constructed from each range image and swept to build the corresponding solid models (Figure 8). The four solids are intersected to produce the final model, shown in wireframe in Figure 9. As can be seen in the final model, there are "boundaries" where the intersection of the solids from two overlapping sensing operations causes an increase in the density of mesh elements. At this level of resolution the model would be a prime candidate for a decimation algorithm such as the one presented in [2]. At this point we have a very reasonable 3-D solid in a CAD format that may be used by other robotics tasks as needed. Refinement of this part may be accomplished using standard CAD primitives. For example, the holes on the sides of the part which were not completely imaged could be introduced by using a through-hole operator present in most CAD packages.

FIGURE 6. Photograph of strut-like part.

7.0 Integrating planning with model acquisition

The ultimate goal of this system is to be able to construct models without any operator intervention. Our current system turns the rotation stage a pre-defined amount between each model acquisition process. While this may be acceptable for some cases, as the objects become more complicated self-occlusion becomes a problem. It becomes increasingly important to have the rotations generated automatically by a planning component which attempts to bring the model to a high level of fidelity while also minimizing the number of additional scans needed. To this end we discuss the integration of our system with a sensor planner to create a fully automated system.

The Sensor Planning problem is that of computing a set of sensor locations for viewing a set of features given a model of a stationary object or scene, a sensor model, and a set of sensing constraints [15]. In our previous work, we've been able to reason effectively about static sensor planning using self occlusion as a guide [14] [1]. This planner is able to reason about self occlusion and compute valid viewpoints that will be free of occlusion given a model feature that needs to be imaged. Hence, as the incremental modeling process proceeds, regions that need more sensing can be guaranteed of having an occlusion free view direction from the sensor.

Consider the U-shaped part shown at the top left of Figure 10. It is easy to see how a scanning process using fixed angles for the turntable rotation will often result in an incomplete model. However, even systems that plan using the volume of occlusion will run into difficulties with objects such as this one that have significant self-occlusion. These systems typically use a histogram of the normals from surfaces of the volume of occlusion, weighted by area, and select the peak to determine the next viewing direction. If a planning system utilizes only information about what is unsensed, and does not incorporate new information learned about the object, it will not be able to identify situations

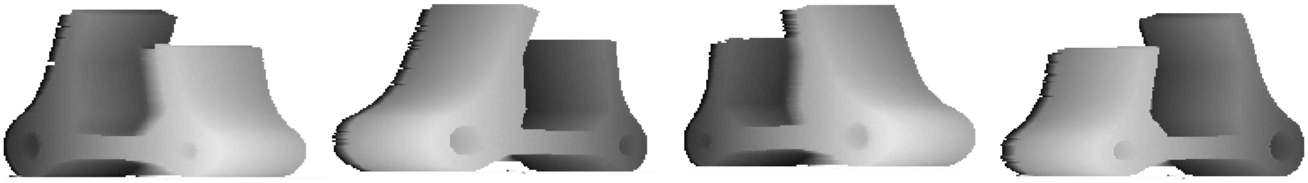

FIGURE 7. Four range images of the part in Figure 6. The turntable rotates 90 degrees between images.

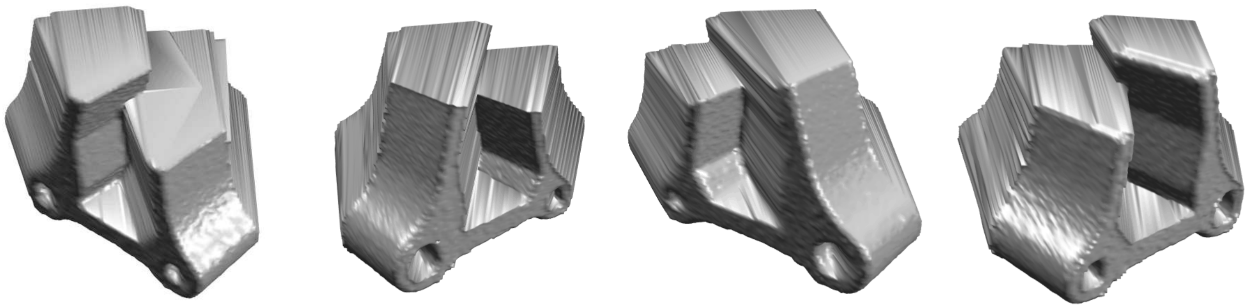

FIGURE 8. Solid models constructed by sweeping the meshes shown in Figure 7.

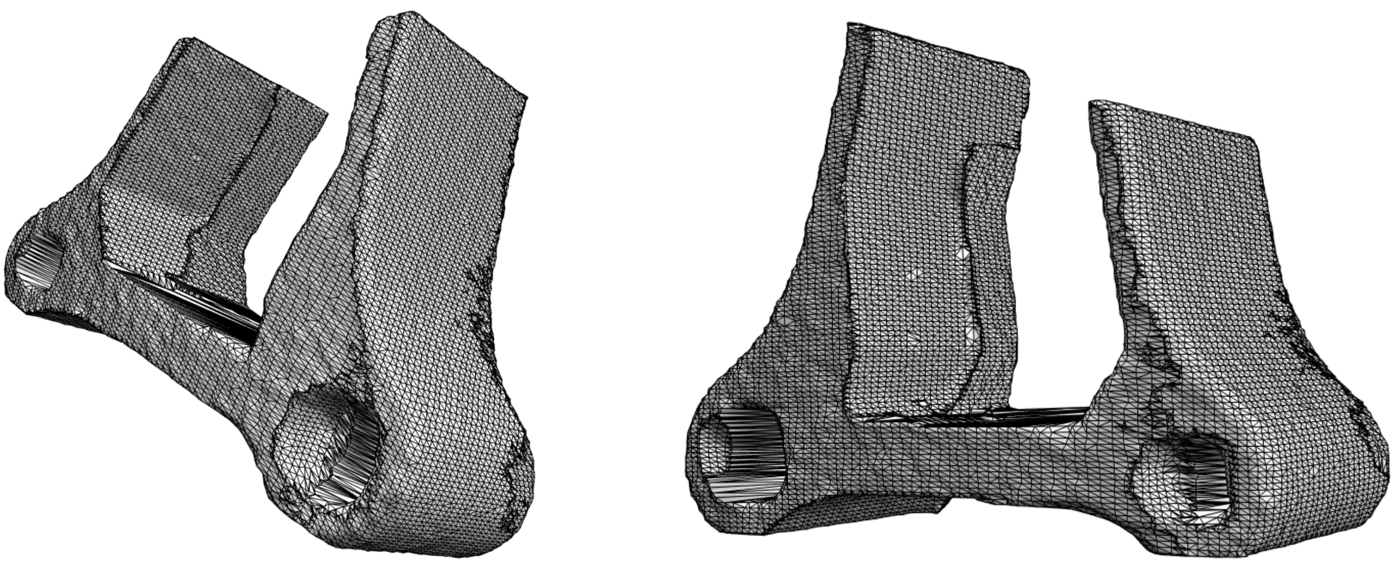

FIGURE 9. Solid formed by intersection of models in Figure 8.

where the object prevents acquisition of some surfaces by self occlusion. In Figure 10 the original part is scanned with an arrow denoting the sensor direction, perpendicular to the dark-bordered face (top left). The large area of occluded volume (shown in grey) on the right side of the first model causes the sensor to image from that direction next (top right). After the model from that viewpoint is integrated, the largest occluded region is the hidden face on the left side, which directs the imaging process to there (lower left). Finally the interior faces become the largest unseen areas. However, without taking into account the model surfaces, the interior face to be imaged will be blocked be one of the "arms" of the "U", shown in cross-hatch in the last model (lower right). The model's interior faces will never be imaged properly

If, however, the model's imaged surfaces are included in the analysis after each addition of new information to determine the next sensing operation, a more complete model may be built as in Figure 11. During the imaging of the object (top left) and of the first two models (top right and second-row left) the behavior is the same as in the previous case, since the surfaces to be imaged are occlusion-free.

However, when attempting to image the dark-bordered interior face (second-row right), the planner detects the possible occlusion, and one of the two non-occluding viewpoints is selected. The resulting model is shown at the bottom of the figure.

We are currently implementing the viewpoint planner that will generate unoccluded sensor positions given a partially built model. This planning component will be able to determine the surfaces in a model which need to be imaged, and either construct a viewpoint that will image some of them or determine that no such viewpoint is possible.

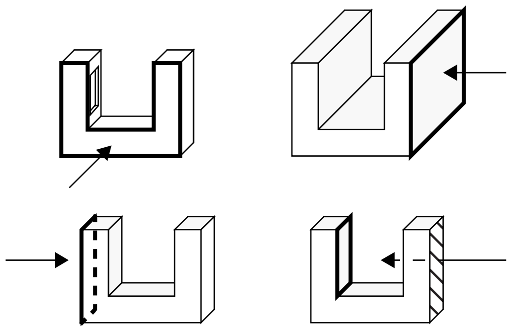

FIGURE 10. Model built with sensor planning relying only on volume of occlusion.

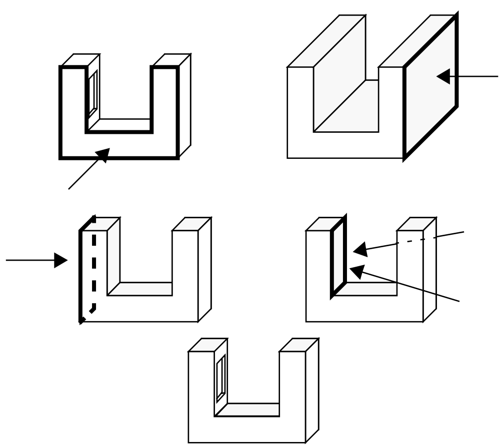

FIGURE 11. Model built with sensor planning based on both the volume of occlusion and the current model.

8.0 Conclusions and future work

In this paper we have presented a method to construct solid models from multiple range images. This method uses modeling techniques from both mesh surface and solid representations. By combining these two we retain the benefits of mesh surfaces, such as representational flexibility and conceptual simplicity, while still allowing the use of well-defined set-theoretic merging operations inherent to solid modelers. This system is able to integrate models from different viewpoints in an incremental fashion that particularly suits the use of on-line planning. Experimental results have been presented for a complex part that includes polygonal faces, curved surfaces, and large self-occlusions. Finally, we have outlined a planning component that will allow us to acquire complete models without user input, while at the same time reducing the number of sensing operations.

9.0 References

[1] Steven Abrams, Peter K. Allen, and Konstantinos A. Tarabanis. Dynamic sensor planning. In *International Conference on Intelligent Autonomous Systems*, pages 206–215, Pittsburgh, PA, February 1993.

[2] J Cohen at al., Simplification Envelopes. In *Proceedings of SIGGRAPH*, pp.119-128, 1996.

[3] C. Connolly. The determination of next best views. In *Proceedings 1985 IEEE International Conference on Robotics and Automation*, pages 432–435, 1985.

[4] C. Connolly and J. Stenstrom. 3d scene reconstruction from multiple intensity images. 1989.

[5] B. Curless and M. Levoy. *A Volumetric Method for Building Complex Models from Range Images*. In *Proceedings of SIGGRAPH*, 1996.

[6] H.Hoppe. *Surface Reconstruction from Unorganized Points*. Ph.D. thesis, Dept. of Com. Science and Engineering, U. of Washington, 1994.

[7] K. Ikeuchi and P.J. Flynn, Editorial: Recent progress in CAD-based vision, Computer VIsion and Image Understanding, 61(3), May 1995.

[8] J. Maver and R. Bajcsy. How to decide from the first view where to look next. In *Proceedings 1990 DARPA Image Understanding Workshop*, pages 482–496, 1990.

[9] W. M. Martin and J. K. Aggarwal. Volumetric descriptions of objects from multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(2):150–158, March 1983.

[10] 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, in *Proceedings SPIE Symposium on Intelligent Systems and Advanced Manufacturing*, Phila. PA, Oct. 1995.

[11] M. Rutishauser, M. Stricker, and M. Trobina. Merging range images of arbitrarily shaped objects. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 573–580, 1994.

[12] G.H. Tarbox and S.N. Gottshlich. Ivis: An integrated volumetric inspection system. *Computer Vision and Image Understanding*, 61(3):430–444, may 1995.

[13] W.B. Thompson, H.J. de St. Germain, T.C. Henderson, and J.C. Owen. Constructing high-precision geometric models from sensed position data. In *Proceedings 1996 ARPA Image Understanding Workshop*, pages 987–994, 1996.

[14] K. Tarabanis, R. Tsai, P.K. Allen. The MVP sensor planning system for robotic tasks. In *IEEE Trans. on Robotics and Automation*, 11(1), February 1995.

[15] K. Tarabanis P.K. Allen, and R. Y. Tsai. A survey of sensor planning in computer vision. In *IEEE Trans. on Robotics and Automation*, 11(1), February 1995.

[16] G. Turk and M. Levoy. Zippered polygon meshes from range images. In *Proceedings of SIGGRAPH*, pp. 311–318, 1994.

[17] P. Whaite and F. Ferrie. Uncertain views. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3–9, 1992.