From Images to Models: Automatic 3D Object Model Construction from Multiple Views

Chitra Dorai[†], Gang Wang[†], Anil K. Jain[†] and Carolyn Mercer[‡]

Department of Computer Science[†]
Michigan State University
East Lansing, MI 48824
{dorai,wanggang,jain}@cps.msu.edu

Optical Technology Branch[‡]
NASA Lewis Research Center
Cleveland, OH 44135
cmercer@lerc.nasa.gov

Abstract

Automatic 3D object model construction is important in applications ranging from manufacturing to entertainment industry, since CAD models of existing objects may be either unavailable or unusable. We describe a prototype system for automatically registering and integrating multiple views of objects using surface depth data to construct their geometric models. New techniques for handling key issues such as robust estimation of view transformations relating multiple views and seamless integration of registered data to form an unbroken surface have been proposed and implemented in our system. Our system has been tested on depth data that were obtained using digital interferometry techniques as well as a laser range scanner. Experimental results on real object surfaces demonstrate the good performance of our system.

1. Introduction

Reverse engineering has emerged as an important concept in modern manufacturing. There is a growing interest in transforming images to models, *i.e.,* to construct geometric and descriptive models of 3D objects from sensed data for applications ranging from component inspection, manufacturing, surgical planning to entertainment industries. Automatic object model construction is useful in duplicating or modifying parts and in verifying the tolerance of manufactured parts for which CAD models may not be available. Accurate models of already existing complex shaped objects are required for synthesizing arbitrary views and also for recognizing them in many applications such as robot navigation, part inspection, animation and visualization in virtual museums and vision augmented environments.

Automatic construction of geometric models of 3D objects involves three major steps: (i) data acquisition, (ii) registration of different views, and (iii) integration. Data acquisition involves obtaining either intensity or depth data of an object from multiple viewpoints. In this paper the term data refers to surface depth data that can be obtained using a laser range scanner or an interferometric sensor. Note that the accurate 3D spatial relations between different views may not be easily and directly obtained in many cases. Therefore, integration of data from multiple views is not only dependent on the representation chosen for the model description, but also requires a knowledge of the transformations relating the data obtained from multiple views. The goal of registration, also known as the *correspondence problem* [3], is to find the transformations that relate multiple views, thus bringing the object regions that are shared between them into alignment. Integration merges data from multiple views using the computed view transformations, to create a single surface representation in a unique coordinate frame.

In this paper, we describe a complete system for registering and integrating multiple views of an object to construct its object model. We have developed a robust approach [5] to register range images of multiple object views in the presence of uncertainties and noise in surface data. It avoids feature detection and computes the relative view transformations directly using the surface depth data. This method also allows data to be obtained from any viewpoint without the need for a control device with six degrees of freedom motion. Our weighted view integration scheme merges the registered views together to result in a continuous surface. Experimental results on real object surfaces demonstrate the good performance of our system.

2. System Overview

A taxonomy and a comprehensive survey of other research efforts [12, 3, 1, 8, 2, 9, 11, 10] directed at building

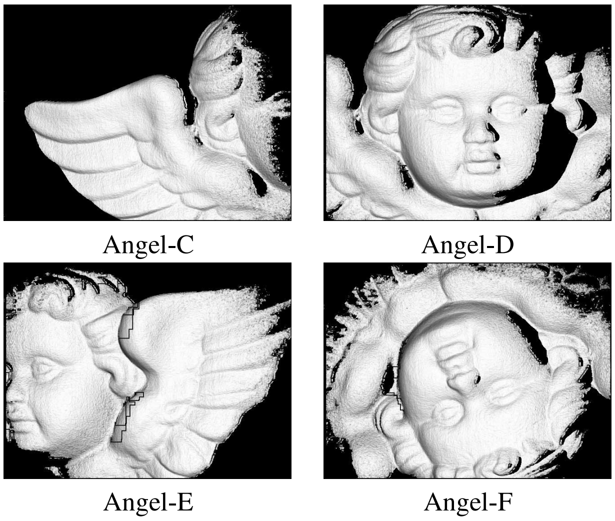

Figure 1. Four different views (range images) of an angel sculpture obtained using a prototype fiber-optic projected fringe system designed and built at NASA Lewis Research Center.

3D object models through registration and integration can be found in [4]. Our multi-view integration system consists of several stages of processing: data acquisition, preprocessing, view registration, and view integration. We describe each processing stage in our model construction system in the sections below.

Surface depth of objects can be measured from multiple viewpoints using a number of techniques such as time-of-flight lasers, structured light, sonar, and radar. Algorithms from computer vision to determine depth from shading, texture, stereo, motion, and focus are also useful to acquire the data. The surface depth (range) information used to create models in this paper is obtained either using structured light (laser range scanners) [6] or using fiber-optic projected-fringe interferometry [7], and the data consists of a discrete set of points present in a regular $X - Y$ grid. The interferometric data was used in our study because it is representative of data encountered in many industrial applications. Figure 1 shows the data from multiple viewpoints obtained using a fiber-optic projected-fringe system [7]. The range data is rendered as pseudo intensity in Figure 1 and points farther from the sensor are shown darker. Projected fringe interferometers are characterized by the tradeoff between high accuracy and measurement area; the width of the measured area can only be up to 2,000 times larger than the depth resolution. Hence for large objects, multiple views are required to get complete surface information.

The surface data obtained from the optical sensors can be noisy. In order to improve the accuracy of registration and data integration, some preprocessing is required to remove the noise in the input images. We have implemented two preprocessing steps: (i) removal of isolated pixels and (ii) removal of noise through median filtering. The experimental results [4] show that the combination of these two preprocessing steps improves the accuracy of integration results noticeably.

2.1. Registration of Views

Given a pair of range images P and Q (different views of the same 3D object), the goal of a registration algorithm is to find the best rigid transformation T that relates P and Q. These two views are said to be *registered* if equation $T\mathbf{p} = \mathbf{q}$ holds for any pair of points $(\mathbf{p}, \mathbf{q})$, $\mathbf{p} \in P$, $\mathbf{q} \in Q$, where $\mathbf{p}$ and $\mathbf{q}$ represent the same surface point. View P is referred to as the *source* image, and view Q is called the *destination* image. An adaptation of the iterative distance minimization technique proposed by Chen and Medioni [3] is employed in our system to register a given pair of object views from their surface data directly.

2.1.1. Pair-wise View Registration

The approach of Chen and Medioni [3] is based on the assumption that an approximate transformation between two views is already known, i.e., data from the two views are approximately registered, and the goal is to refine the initial estimate to obtain a more accurate global registration. Nearly all distance minimization algorithms suppose that a coarse initial estimate of the transformation is either available from the position of the sensor or estimated using other feature-based techniques. Given a set of N pairs of *corresponding* points, called *control points* in two views, $\mathbf{p}_i \in P$ and $\mathbf{q}_i \in Q$, $\quad i = 1 \ldots N$, where $N \geq 3$, the transformation $T = (\alpha, \beta, \gamma, t_x, t_y, t_z)$ can be estimated by minimizing

$$e = \sum_{i=1}^{N} ||T\mathbf{p}_i - \mathbf{q}_i)||^2. \tag{1}$$

Since $\mathbf{q}_i \in Q$ *corresponding* to a point $\mathbf{p}_i \in P$ is not usually known, Chen and Medioni [3] used the following objective function to iteratively minimize the distances from points on one surface to another:

$$e^k = \sum_{i=1}^{N} d_s^2(T^k\mathbf{p}_i, S_i^k), \tag{2}$$

where T^k is the 3D transformation applied to a control point $\mathbf{p}_i \in P$ at the kth iteration, $l_i = \{\mathbf{a} \mid (\mathbf{p}_i - \mathbf{a}) \times \mathbf{n}_{\mathbf{p}_i} = 0\}$ is the line normal to P at $\mathbf{p}_i$, $\mathbf{q}_i^k = (T^k l_i) \cap Q$ is the intersection point of surface Q with the transformed line $T^k l_i$, $\mathbf{n}_{\mathbf{q}_i}^k$

is the normal to Q at $\mathbf{q}_i^k$, $S_i^k = \{\mathbf{s} \mid \mathbf{n}_{\mathbf{q}_i}^k \cdot (\mathbf{q}_i^k - \mathbf{s}) = 0\}$ is the tangent plane to Q at $\mathbf{q}_i^k$ ('$\cdot$' stands for the scalar product and '$\times$' for the vector product), and d_s is the signed distance from a 3D point to a plane. The tangent plane S_i^k serves as a local linear approximation to the surface Q at point $\mathbf{q}_i$. The intersection point $\mathbf{q}_i^k$ is an approximation to the actual *corresponding* point $\mathbf{q}_i$ that is unknown at each iteration k. The registration algorithm thus, finds a transformation T that iteratively minimizes e^k, using a least squares method [4].

2.1.2. Verification of Control Point Correspondences

Given a set of control points in the *source* range image to be registered, finding their *corresponding* control points in the *destination* range image is one of the primary issues in the registration algorithm. The effectiveness of the algorithm for establishing the correspondence depends upon the following two factors: (i) the accuracy of the initial transformation that drives the iterative process and (ii) the smoothness of the surfaces to be registered. With discrete range data, if the initial guess is accurate and the object surfaces are smooth, most corresponding control points found by performing an iterative search near the neighborhood of prospective intersection points in the *destination* image are reasonable approximations to the true corresponding control points. However, if the initial guess is not accurate, or the surface of the object is noisy and rough, many corresponding control points found may be far away from the actual corresponding control points. Any inaccuracy in establishing the correspondence of control points leads to inaccurate estimation of d_s and hence, the transformation parameters. We have designed a verification mechanism to check the validity of the corresponding control points found by the algorithm, and only valid corresponding control point pairs are used to update the transformation matrix.

The correctness of the correspondences of control points is established by exploiting the distance constraint imposed by rigid transformations. Given control points $\mathbf{p}_1$ and $\mathbf{p}_2$ on the surface P, and the corresponding control points $\mathbf{q}_1$ and $\mathbf{q}_2$ on the surface Q, the constraint $||\mathbf{p}_1 - \mathbf{p}_2|| = ||\mathbf{q}_1 - \mathbf{q}_2||$ holds for all rigid transformations. Otherwise, $(\mathbf{p}_1, \mathbf{q}_1)$ and $(\mathbf{p}_2, \mathbf{q}_2)$ cannot be valid control point pairs at the same time and thus, $(\mathbf{p}_1, \mathbf{q}_1)$ and $(\mathbf{p}_2, \mathbf{q}_2)$ are not compatible with each other. Two pairs of control points $(\mathbf{p}_1, \mathbf{q}_1)$ and $(\mathbf{p}_2, \mathbf{q}_2)$ are defined to be *compatible* if and only if

$$-t \leq \frac{||\mathbf{p}_1 - \mathbf{p}_2|| - ||\mathbf{q}_1 - \mathbf{q}_2||}{\max(||\mathbf{p}_1 - \mathbf{p}_2||, ||\mathbf{q}_1 - \mathbf{q}_2||)} \leq t, \tag{3}$$

where t is a pre-defined threshold (10% in our implementation [4]). Compared with the registration algorithm of Chen and Medioni [3], the improved algorithm with the control point verification step is found to provide more accurate results as demonstrated by our experiments.

2.2. Multiple View Integration

Given a set of registered views, one of the goals in designing an integration algorithm is that the assembled data should be as smooth as possible without any loss of details present in the original raw data. Given the fact that there can be some imprecision in the estimated view transformation parameters obtained in the registration stage, and that the input data may be noisy, it is not quite easy to achieve this goal. Although a smoothing algorithm appears to be necessary to smooth the integrated data, surface details may be lost during blurring. We have proposed and implemented a weighted integration algorithm to merge a set of registered range images seamlessly.

Given multiple views $\{S_1, S_2, \cdots S_n\}$ and the corresponding transformation matrices $\{T_1, T_2, \cdots T_n\}$ that relate these views to one another, a 3D model is constructed as a 3D point set in the following manner:

$$M_1 = \bigcup_{i=1}^{n} (T_i \circ S_i) \tag{4}$$

in which S_i is the set of 3D points obtained from the range image of the ith view. However, this method has a severe drawback. Consider points $\{\mathbf{p}_{i_1}, \mathbf{p}_{i_2}, \cdots \mathbf{p}_{i_k}\}$, $\mathbf{p}_{i_j} \in S_j$, $j = 1, 2, \cdots k$ that represent the same point $\mathbf{p}_i$ on the object surface visible in multiple object views, since the transformation matrix T_i estimated with the registration stage may not be very accurate. I.e., points $T_j \circ \mathbf{p}_{i_j}$, $j = 1, 2, \cdots k$ may not reside exactly at the same 3D location. Therefore, inconsistencies may be introduced in the resultant surface model. To avoid this problem, we decompose the cartesian 3D space into multiple equal-sized vessels $\{V_{ijk}\}$. For a given vessel V_{ijk}, we obtain a single 3D point $\mathbf{p}_{ijk}$ that represents all those points that fall into V_{ijk}:

$$\mathbf{p}_{ijk} = \frac{\sum_{\forall \mathbf{p} \in S\mathbf{p}_{ijk}} W(\mathbf{p}) \times \mathbf{p}}{\sum_{\forall \mathbf{p} \in S\mathbf{p}_{ijk}} W(\mathbf{p})} \tag{5}$$

where $S\mathbf{p}_{ijk} = \{\mathbf{p} \mid \mathbf{p} \in M_1, \mathbf{p} \in V_{ijk}\}$ and $W(\mathbf{p})$ is the weight associated with point $\mathbf{p}$. Note that each point (pixel) in an input range image is associated with a weight, which is essentially the distance of the pixel to the nearest boundary. The resultant surface model is $M_w = \{\mathbf{p}_{ijk}\}$.

Once the 3D model M_w of a given object is obtained, a user can visualize it from various new viewpoints. For a given viewpoint, there exists an equivalent transformation $R = \{\alpha, \beta, \gamma, t_x, t_y, t_z\}$, based on which a rigid transformation matrix T can be computed. The visible point subset $V_{set} \subseteq T \circ M_w$ can be determined based on the depth value of each point in $T \circ M_w$.

3. Experimental Results

We tested the performance of our prototype system on surface data obtained using a Technical Arts White laser range scanner available at Michigan State University and a digital interferometric sensor available at NASA Lewis Research Center. We registered and integrated the angel images shown in Figure 1 successfully (see Figure 2). In order to compare the registration accuracy of our algorithm with the registration algorithm [3] proposed by Chen and Medioni, we used the same input images. The initial guess of the transformation relating a pair of views is set by visually inspecting the range images. When the range images contain the entire object surface and the rotations of the object in the views are primarily in the plane, then a good initial guess for the iterative algorithm can be determined automatically by computing the approximate rotation and translation from the major (principal) axes of the object computed using the range data [5].

Figure 2. Result of merging all angel views.

In all our experiments, we used registration error (δ_r) as an indication of the accuracy of registration. It was computed as the average distance from the surface of the *source* image to the *destination* image in the direction of the Z axis, once the *source* image has been transformed using the transformation parameters determined by the registration process. In most cases, only parts of the *source* and *destination* images are overlapped. Therefore, δ_r is calculated only in the overlapping regions. δ_r is primarily determined by two factors: (i) the accuracy of registration and (ii) the noise in the input images. Though it is hard to quantify the contribution of each factor to the final value of δ_r, it can still be used to compare the relative merits of different registration techniques, since the noise in the input images is the same. In the case of angel images, the distance between the two eyes is about 1.2 cms (140 pixels) and the range image of Angel-D spans about 1.4 cms of depth. The error in registering Angel-C and Angel-D is about 3% of the depth span (about 4 pixels).

Results from testing our system also on Big-Y (a pipe object) whose images were obtained using the White scanner indicate that accurate registration results from our improved registration method lead to better integration of data especially when the input images are complex and noisy [4]. These tests also show that our method requires fewer number of iterations to achieve convergence. Given a coarse correct initial guess, our system usually takes about 30 seconds to register two range images whose sizes are 640×480 on a SPARCstation 10 with 32MB RAM.

By registering and integrating nine different views of a wooden Buddha sculpture shown in Figure 3, a 3D surface model was constructed which allows a user to rotate and view the Buddha from new viewing directions (see Figure 4) distinct from those that were used to generate the input. The thickness of the laser beam generated by the range scanner is about 3mm. As a result, many of the fine details on the head of the Buddha were lost or were not measured accurately while acquiring the data using the Technical Art White scanner. This loss of details cannot be properly compensated during the construction of the 3D surface model. Thus, on the resultant 3D surface model, there are a few small "holes" on the Buddha's head. The input views contain altogether 142,448 3D surface points and each pixel in an image represents an actual size of $0.026'' \times 0.045''$. The vessel size selected was $0.04''$ during view integration. The resultant surface model contains 47,720 3D points. Our experimental results demonstrate that our system is capable of seamlessly merging surface information from multiple range images, with very small noticeable loss of details. We believe that our system is of practical value for many applications due to its accuracy and robustness.

4. Summary

We have proposed and implemented a complete prototype system for registering and integrating multiple 3D object views to construct their geometric models from range data. We have employed an improved registration technique to compute the view transformation relating a pair of views. Our experimental results illustrate that by combining preprocessing techniques such as removal of isolated pixels and median filtering, and by enhancing the performance of the registration algorithm with the verification of control point correspondences, the accuracy of view registration results can be improved noticeably. A weighted integration algorithm which results in smoothly merged surfaces is used to construct a 3D model. We have demonstrated that it is feasible to construct the surface model of a complex shaped 3D object from its multiple views accurately.

References

[1] P. J. Besl and N. D. Mckay. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine*

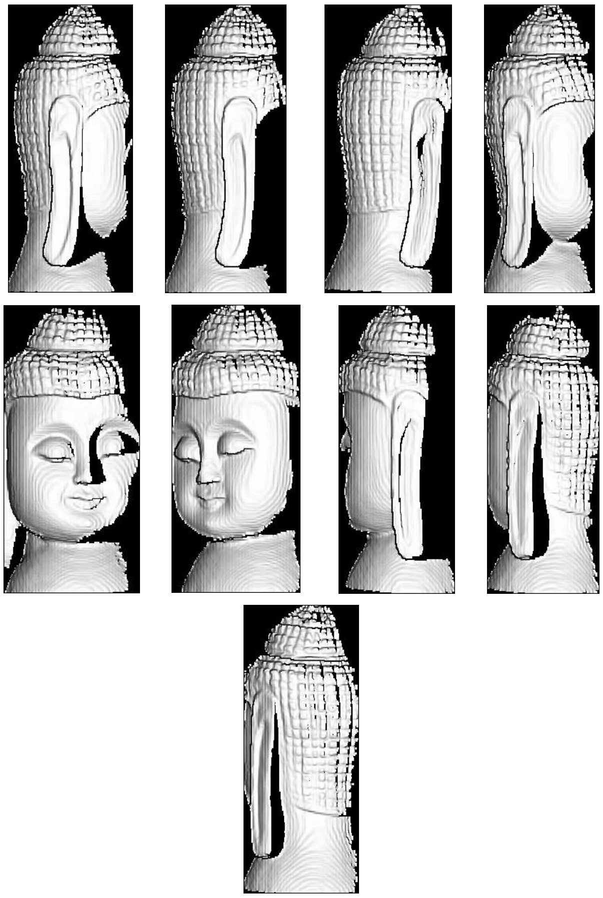

Figure 3. Nine views used to reconstruct a 3D surface model of the Buddha.

Intelligence, 14(2):239–256, 1992.

[2] G. Blais and M. D. Levine. Registering multiview range data to create 3d computer graphics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):820–824, 1995.

[3] Y. Chen and G. Medioni. Object modelling by registration of multiple range images. *Image and Vision Computing*, 10(3):145–155, April 1992.

[4] C. Dorai, G. Wang, A. K. Jain, and C. Mercer. From images to models: automatic 3d object model construction from multiple views. Technical report, Department of Computer Science, Michigan State University, East Lansing, January 1996.

[5] C. Dorai, J. Weng, and A. K. Jain. Optimal registration of multiple range views. In *Proc. 12th International Conference on Pattern Recognition*, pages 569–571, Jerusalem, Israel, October 1994.

[6] R. Jarvis. Range sensing for computer vision. In A. K. Jain and P. J. Flynn, editors, *Three-Dimensional Object Recognition Systems*, pages 17–56. Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1993.

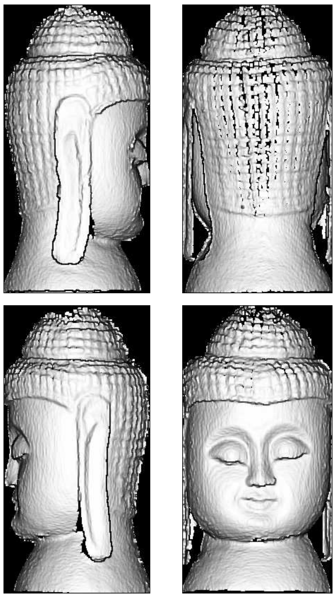

Figure 4. Four new views generated using the reconstructed surface model of the Buddha.

[7] C. R. Mercer and G. Beheim. Fiber-optic projected-fringe digital interferometry. In *Hologram Interferometry and Speckle Metrology Proceedings*, pages 210–216, Bethel, CT, 1990. Society for Experimental Mechanics.

[8] M. Potmesil. Generating models for solid objects by matching 3D surface segments. In *Proc. International Joint Conference on Artificial Intelligence*, pages 1089–1093, Karlsruhe, Germany, 1983.

[9] M. Soucy and D. Laurendeau. A general surface appraoch to the integration of a set of range views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(4):344–358, 1995.

[10] W. B. Thompson, J. C. Owen, and H. J. de St. Germain. Feature-based reverse engineering of mechanical parts. Technical Report UUCS-95-010, Department of Computer Science, University of Utah, Salt Lake City, Utah, 1995.

[11] G. Turk and M. Levoy. Zippered polygon meshes from range images. In *SIGGRAPH 94*, 1994.

[12] B. C. Vemuri and J. K. Aggarwal. 3-D model construction from multiple views using range and intensity data. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 435–437, Miami Beach, FL, 1986.