# Short Papers

## Optimal Registration of Object Views Using Range Data

Chitra Dorai, *Member, IEEE*, John Weng, *Member, IEEE*, and Anil K. Jain, *Fellow, IEEE*

**Abstract**—This paper deals with robust registration of object views in the presence of uncertainties and noise in depth data. Errors in registration of multiple views of a 3D object severely affect view integration during automatic construction of object models. We derive a minimum variance estimator (MVE) for computing the view transformation parameters accurately from range data of two views of a 3D object. The results of our experiments show that view transformation estimates obtained using MVE are significantly more accurate than those computed with an unweighted error criterion for registration.

**Index Terms**—Image registration, view transformation estimation, view integration, automatic object modeling, 3D free-form objects, range data.

### 1 Introduction

AN important issue in the design of 3D object recognition systems is building models of physical objects. Object models are extensively used for synthesizing and predicting object appearances from desired viewpoints, and also for recognizing them in many applications, such as robot navigation and industrial inspection. It becomes necessary on many occasions to construct models from multiple measurements of 3D objects, especially when a precise geometric model such as a CAD description is not available and cannot be easily obtained. This need is felt particularly with 3D free-form objects, such as sculptures and human faces, that may not possess simple analytical shapes for representation. With growing interest in creating virtual museums and virtual reality functions such as walkthroughs, creating computer images corresponding to arbitrary views of 3D scenes and objects remains a challenge.

Automatic construction of 3D object models typically involves three steps:

1) data acquisition from multiple viewpoints,
2) registration of views, and
3) integration.

Data acquisition involves obtaining either intensity or depth data of multiple views of an object. Integration of multiple views is dependent on the representation chosen for the model and requires knowledge of the transformation relating the data obtained from different viewpoints. The intermediate step, registration, is also known as the *correspondence problem* [1], and its goal is to find the transformations that relate the views. Inaccurate registration leads to greater difficulty in seamlessly integrating the data. It ultimately affects surface classification, since surface patches from different views may be erroneously merged, resulting in holes and discontinuities in the merged surface. For smooth merging of data, accurate estimates of transformations are vital. In this paper, we focus on the issue of pairwise registration of noisy range images of an object obtained from multiple viewpoints using a laser range scanner.

We derive a minimum variance estimator to compute the transformation parameters accurately from range data. We investigate the effect of surface measurement noise on the *registration* of a pair of views, and propose a new method that improves upon the approach of Chen and Medioni [1]. We have not seen any work that reports, to date, establishing the dependencies between the orientation of a surface, noise in the sensed surface data, and the accuracy of surface normal estimation, and how these dependencies can affect the estimation of 3D transformation parameters that relate a pair of object views. We present a detailed analysis of this "orientation effect" with geometrical arguments and experimental results.

### 2 Previous Work

There have been several research efforts directed at solving the registration problem. While the first category of approaches relies on precisely calibrated data acquisition devices to determine the transformations that relate the views, the second kind involves techniques to estimate the transformations from the data directly. The calibration-based techniques are inadequate for constructing a complete description of complex-shaped objects, as views are restricted to rotations or to some known viewpoints only, and, therefore, the object surface geometry cannot be exploited in the selection of vantage views to obtain measurements.

With the second kind, interimage correspondence has been established by matching the data or the surface features derived from the data [2]. The accuracy of the feature detection method employed determines the accuracy of feature correspondences. Potmesil [3] matched multiple range views using a heuristic search in the view transformation space. Though quite general, this technique involves searching a huge parameter space, and, even with good heuristics, it may be computationally very expensive. Chen and Medioni avoid the search by assuming an initial approximate transformation for the registration, which is improved with an iterative algorithm [1] that minimizes the distance from points in a view to tangential planes at corresponding points in other views. Besl and McKay [4], Turk and Levoy [5], and Zhang [6] employ variants of the iterated closest-point algorithm. Blais and Levine [7] propose a reverse calibration of the range-finder to determine the point correspondences between the views directly, and use stochastic search to estimate the transformation. These approaches, however, do not take into account the presence of noise or inaccuracies in the data and its effect on the estimated view-transformation. Our registration technique also uses a distance minimization algorithm to register a pair of views, but we do not impose the requirement that one surface has to be strictly a subset of the other. While our approach studies, in detail, the effect of noise on the objective function [1] that is being minimized and proposes an improved function to register a pair of views, Bergevin et al. [8], [9] propose to register all views simultaneously to avoid error accumulation due to sequential registration. Haralick et al. [10] have also showed that a weighted least-squares technique is robust under noisy conditions under various scenarios, such as 2D-2D, 3D-3D image registration.

### 3 A Nonoptimal Algorithm for Registration

Two views, $P$ and $Q$, of a surface are said to be registered when *any* pair of points, $\mathbf{p}$ and $\mathbf{q}$, from the two views representing the same

- *C. Dorai is with the IBM T.J. Watson Research Center, P.O. Box 704, Yorktown Heights, NY 10598. E-mail: dorai@watson.ibm.com.*
- *J. Weng and A.K. Jain are with the Department of Computer Science, Michigan State University, East Lansing, MI 48824.*
- *E-mail: {weng, jain}@cps.msu.edu.*

*Manuscript received 12 Aug. 1996. Recommended for acceptance by K.Bowyer.*
*For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 104916.*

object surface point can be related to each other by a *single* rigid 3D spatial transformation $\mathcal{T}$, so that $\forall \mathbf{p} \in P, \exists \mathbf{q} \in Q$ such that $\| \mathcal{T}\mathbf{p} - \mathbf{q} \| = 0$, where $\mathcal{T}\mathbf{p}$ is obtained by applying the transformation $\mathcal{T}$ to $\mathbf{p}$, and $\mathcal{T}$ is expressed in homogeneous coordinates as a function of three rotation angles, $\alpha$, $\beta$, and $\gamma$ about the $x$, $y$, and $z$ axes, respectively, and three translation parameters, $t_x$, $t_y$, and $t_z$. The terms "view" and "image" are used interchangeably in this paper. The approach of [1] is based on the assumption that an approximate transformation between two views is already known, and the goal is to refine the initial estimate to obtain more accurate global registration. The following objective function was used to minimize the distances from surface points in one view to another iteratively:

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

where $\mathcal{T}^k$ is the 3D transformation applied to a control point $\mathbf{p}_i \in P$, $i = 1 \ldots N$ at the $k$th 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 $\mathbf{p}_i$, $\mathbf{q}_i^k = (\mathcal{T}^k l_i) \cap Q$ is the intersection point of surface $Q$ with the transformed line $\mathcal{T}^k l_i$, $\mathbf{n}_{\mathbf{q}_i}^k$ 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 $\mathbf{q}_i^k$, and $d_s$ is the signed distance from a point to a plane as given in (2). Note that "$\cdot$" stands for the scalar product and "$\times$" for the vector product. Fig. 1 illustrates the distance measure $d_s$ between surfaces $P$ and $Q$.

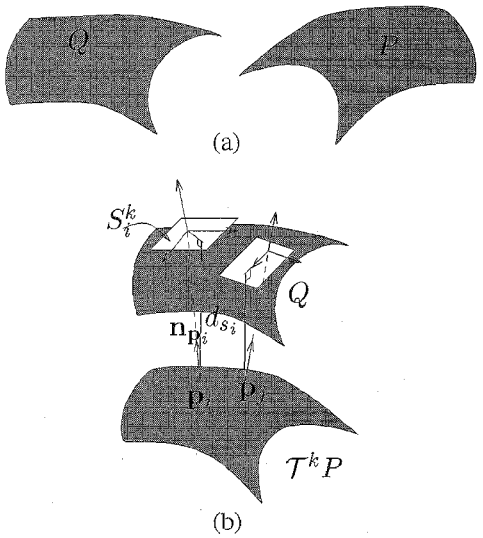



Fig. 1. Point-to-plane distance: (a) Surfaces $P$ and $Q$ before the transformation $\mathcal{T}^k$ at iteration $k$ is applied; (b) distance from the point $\mathbf{p}_i$ to the tangent plane $S_i^k$ of $Q$.

This registration algorithm thus finds a $\mathcal{T}$ that minimizes $e^k$, using a least-squares method iteratively. The tangent plane $S_i^k$ serves as a local linear approximation to the surface $Q$ at a point. The intersection point $\mathbf{q}_i^k$ is an approximation to the actual *corresponding* point $\mathbf{q}_i$ that is unknown at each iteration $k$. An initial $\mathcal{T}^0$ that approximately registers the two views is used to start the iterative process. The signed distance $d_s$, from a transformed point $\mathcal{T}\mathbf{p}_i$, $\mathbf{p}_i \in P$ to a tangential plane $S_i^k \in Q$ is given by

$$d_s = -\frac{\mathcal{A}x + \mathcal{B}y + Cz + \mathcal{D}}{\sqrt{\mathcal{A}^2 + \mathcal{B}^2 + C^2}}, \tag{2}$$

where $\mathcal{T}\mathbf{p}_i = (x, y, z)^T$ and $S_i^k = (\mathcal{A}, \mathcal{B}, C, \mathcal{D})^T$ define the transformed point and the tangential plane, respectively. Note that $(x, y, z)^T$ is the transpose of the vector $(x, y, z)$. By minimizing the distance from a point to a plane, only the direction in which the distance can be reduced is constrained. The convergence of the process can be tested by verifying that the difference between the errors $e^k$ at any two consecutive iterations is less than a prespecified threshold. The line-surface intersection given by the intersection of the normal line $l_i$ and $Q$ is found using an iterative search near the neighborhood of prospective intersection points.

## 4 Registration and Surface Error Modeling

Range data are often corrupted by measurement errors and, sometimes, lack of data. The errors in surface measurements of an object include scanner errors, camera distortion, and spatial quantization, and the missing data can be due to self-occlusion or sensor shadows. Due to noise, it is generally impossible to obtain a solution for a rigid transformation that fits two sets of noisy three-dimensional points exactly. The least-squares solution in [1] is nonoptimal, as it does not handle the errors in $z$ measurements, and it treats all surface measurements with different reliabilities equally. Our objective is to derive a transformation that globally registers the noisy data in some optimal sense. With range sensors that provide measurements in the form of a graph surface $z = f(x, y)$, it is assumed that the error is present along the $z$ axis only, as the $x$ and $y$ measurements are usually laid out in a grid. There are different uncertainties along different surface orientations, and they need to be handled appropriately during view registration. Furthermore, the measurement error is not uniformly distributed over the entire image. The error may depend on the position of a point, relative to the object surface. A measurement error model dealing with the sensor's viewpoint has been previously proposed [11] for surface reconstruction, where the emphasis was recovering straight line segments from noisy single-scan 3D surface profiles.

In this paper, we show that the noise in $z$ values affects the estimation of the tangential plane parameters differently depending on how the surface is oriented. Since the estimated tangential plane parameters play a crucial role in determining the distance $d_s$ which is being minimized to estimate $\mathcal{T}$, we study the effect of noise on the estimation of the parameters of the plane fitted and on the minimization of $d_s$. The error in the iterative estimation of $\mathcal{T}$ is a combined result of errors in each control point $(x, y, z)^T$ from view 1 and errors in fitting tangential planes at the *corresponding* control points in view 2.

### 4.1 Fitting Planes to Surface Data with Noise

Fig. 2 illustrates the effect of noise in the values of $z$ on the estimated plane parameters. For the horizontal plane shown in Fig. 2a, an error in $z$ (the uncertainty region around $z$) directly affects the estimated surface normal. In the case of an inclined plane, the effect of errors in $z$ on the surface normal to the plane is much less pronounced as shown in Fig. 2b. Here, even if the error in $z$ is large, only its projected error along the normal to the plane affects the normal estimation. This projected error becomes smaller than the actual error in $z$ as the normal becomes more and more inclined with respect to the vertical axis. Therefore, our hypothesis is that, as the angle between the vertical (Z) axis and the normal to the plane increases, the difference between the fitted plane parameters and the actual plane parameters should decrease.

We carried out simulations to study the actual effect of the noise in the $z$ measurements on estimating the plane parameters and to verify our hypothesis. The conventional method for fitting planes to a set of 3D points uses a linear least-squares algorithm. This linear regression method implicitly assumes that two of the

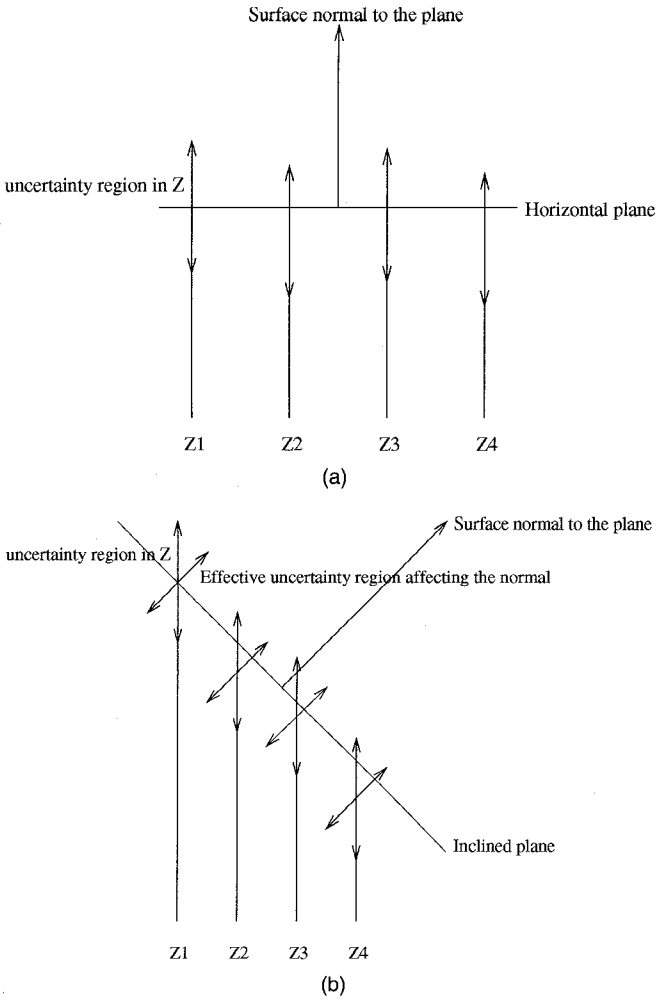



Fig. 2. Effect of noise in $z$ measurements on the fitted normal: (a) when the plane is horizontal; (b) when it is inclined. The double-headed arrows indicate the uncertainty in depth measurements.

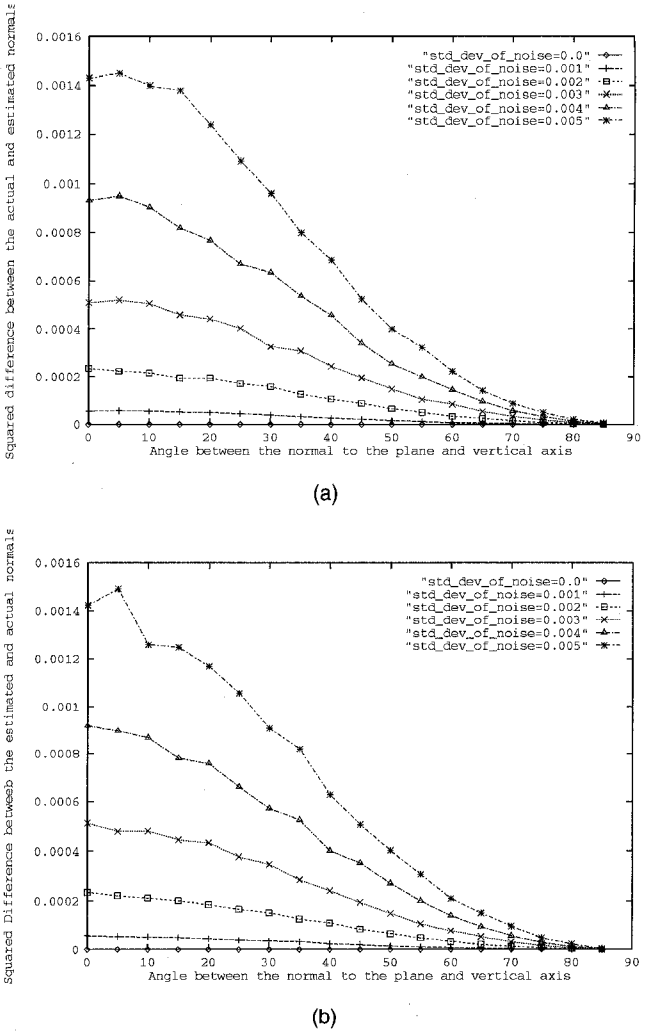



Fig. 3. Effect of *i.i.d.* Gaussian noise in $z$ measurements on the plane: (a) estimated using eigenvector approach; (b) estimated using linear regression.

three coordinates are measured without errors. However, it is possible that, in general, surface points can have errors in all three coordinates, and surfaces can be in any orientation. Hence, we used a classical eigenvector method (principal components analysis) [12] that allows us to extract all linear dependencies.

Let the plane equation be $\mathcal{A}x + \mathcal{B}y + Cz + \mathcal{D} = 0$. Let $X_i = (x_i, y_i, z_i)^T$, $i = 1, 2, \cdots, n$, be a set of surface measurements used in fitting a plane at a point on a surface. Let

$$A = \begin{bmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & z_n & 1 \end{bmatrix} \tag{3}$$

and $h = (\mathcal{A}, \mathcal{B}, C, \mathcal{D})^T$ be the vector containing the plane parameters. We solve for the vector $h$, such that $\|Ah\|$ is minimized. The solution of $h$ is a unit eigenvector of $A^TA$ associated with the smallest eigenvalue. We renormalize $h$, such that $(\mathcal{A}, \mathcal{B}, C)^T$ is the unit normal to the fitted plane, and $\mathcal{D}$ is the distance of the plane from the origin of the coordinate system. This planar fit minimizes the sum of the squared perpendicular distances between the data points and the fitted plane, and is independent of the choice of the coordinate frame.

In our computer simulations, we used synthetic planar patches as test surfaces. The simulation data consisted of surface measurements from planar surfaces at various orientations with respect to the vertical axis. Independent and identically distributed (*i.i.d.*) Gaussian and uniform noise with zero mean and different variances were added to the $z$ values of the synthetic planar data. The standard deviation of the noise used was in the range 0.001-0.005 *in.*, as this realistically models the error in $z$ introduced by a Technical Arts 100X range scanner [13] that was employed to obtain the range data for our experiments. The planar parameters were estimated using the eigenvector method at different surface points, with a neighborhood of size $5 \times 5$. The error $E_{fit}$ in fitting the plane was defined as the norm of the difference between the actual normal to the plane and the normal of the fitted plane estimated with the eigenvector method. Fig. 3a shows the plot of $E_{fit}$ versus the orientation (with respect to the vertical axis) of the normal to the simulated plane at different noise variances. The plot shows $E_{fit}$ averaged over 1,000 trials at each orientation.

It can be seen from Fig. 3a that, in accordance with our hypothesis, the error in fitting a plane decreases with an increase in the angle between the vertical axis and the normal to the plane. When the plane is nearly horizontal (i.e., the angle is small), the error in $z$ entirely contributes to $E_{fit}$ as indicated by Fig. 2a. The error plots for varying amounts of variance were observed to have the same behavior with orientation as shown in Fig. 3a. Similar curves were also obtained with a uniform noise model [14]. These simulations confirm our hypothesis about the effect of noise in $z$ on the fitted plane parameters as the surface orientation changes.

We repeated the simulations using the linear regression method to fit planes to surface data. We refer the reader to [14] for details. Fig. 3b shows the error $E_{fit}$ between the fitted and actual normals to the plane at various surface orientations when *i.i.d.* Gaussian noise was added to the $z$ values. Our hypothesis is well supported by this error plot also.

### 4.2 Proposed Optimal Registration Algorithm

Since the estimated tangential plane parameters are affected by the noise in $z$ measurements, any inaccuracies in the estimates, in turn, influence the accuracy of the estimates of $d_s$, thus affecting the error function being minimized during the registration. Further, errors in $z$ themselves affect $d_s$ estimates (see (2)). Therefore, we characterize the error in the estimates of $d_s$ by modeling the uncertainties associated with them using weights. Our approach is inspired by the Gauss-Markov theorem [15], which states that an unbiased linear minimum variance estimator of a parameter vector $\mathbf{m}$ when $\mathbf{y} = \mathbf{f}(\mathbf{m}) + \delta_y$ is the one that minimizes $(\mathbf{y} - \mathbf{f}(\mathbf{m}))^T \Gamma_y^{-1} (\mathbf{y} - \mathbf{f}(\mathbf{m}))$, where $\delta_y$ is a random noise vector with zero mean and covariance matrix $\Gamma_y$. Based on this theorem, we formulate an optimal error function for registration of two object views as

$$e^k = \sum_{i=1}^{N} \frac{1}{\sigma_{d_s}^2} d_s^2\left(\mathcal{T}^k \mathbf{p}_i, S_i^k\right), \tag{4}$$

where $\sigma_{d_s}^2$ is the estimated variance of the distance $d_s$. When the reliability of a $z$ value is low, the variance of the distance $\sigma_{d_s}^2$ is large, and the contribution of $d_s$ to the error function is small, and when the reliability of the $z$ measurement is high, $\sigma_{d_s}^2$ is small, and the contribution of $d_s$ is large; $d_s$ with a minimum variance affects the error function more. One of the advantages of this minimum variance criterion is that we do not need the exact noise distribution. We only require that the noise distribution be well-behaved and have short tails. In our simulations, we employ both Gaussian and uniform noise distributions to illustrate the effectiveness of our method. We need to know only the second-order statistics of the noise distribution, which, in practice, can often be estimated.

### 4.3 Estimation of the Variance $\sigma_{d_s}^2$

We need to estimate $\sigma_{d_s}^2$ to model the reliability of the computed $d_s$ at each control point, which can then be used in our optimal error function in (4). Let the set of all the surface points be denoted by $P$, and the errors in the measurements of these points be denoted by a random vector $\epsilon$. The error $e_{d_s}$ in the distance computed is due to the error in the estimated plane parameters and the error in the $z$ measurement, and, therefore, is a function of $P$ and $\epsilon$. Since we do not know $\epsilon$, if we can estimate the standard deviation of $e_{d_s}$ (with $\epsilon$ as a random vector) from the noise-corrupted surface measurements $P$, we can use it in (4).

#### 4.3.1 Estimation of $\sigma_{d_s}^2$ Based on Perturbation Analysis

Perturbation analysis is a general method for analyzing the effect of noise in data on the eigenvectors obtained from the data. It is general, in the sense that errors in $x$, $y$, and $z$ can all be handled. This analysis is also related to the general eigenvector method that we studied for plane estimation. The analysis for estimating $\sigma_{d_s}^2$ is simpler if we use linear regression method to do plane fitting [14].

Since we fit a plane with the eigenvector method that uses the symmetric matrix $C = A^T A$ computed from the $(x, y, z)$ measurements in the neighborhood of a surface point, we need to analyze how a small perturbation in the matrix $C$ caused by the noise in the measurements can affect the eigenvectors. Recall that these eigenvectors determine the plane parameters $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})^T$, which, in turn, determine the distance $d_s$. We assume that the noise in the measurements has zero mean and some variance, and that the latter can be estimated empirically. The correlation in noise at different points is assumed to be negligible. Estimation of correlation in noise is very difficult, but, even if we estimate it, its impact may turn out to be insignificant. We estimate the standard deviation of errors in the plane parameters and in $d_s$ on the basis of the first-order perturbations, i.e., we estimate the "linear terms" of the errors.

Before we proceed, we discuss some of the notational conventions that are used: $I_m$ is an $m \times m$ identity matrix; $diag(a, b)$ is a $2 \times 2$ diagonal matrix with $a$ and $b$ as its diagonal elements. Given a noise-free matrix $A$, its noise-matrix is denoted by $\Delta_A$, and the noise-corrupted version of $A$ is denoted by $A(\epsilon) = A + \Delta_A$. The vector $\delta$ is used to indicate the noise vector, $X(\epsilon) = X + \delta_X$. We use $\Gamma$ with a corresponding subscript to specify the covariance matrix of the noise vector/matrix. For a given matrix $A = [A_1\ A_2 \cdots A_n]$, a vector $\mathbf{A}$ can be associated with it as

$$\mathbf{A} = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix}$$

$\mathbf{A}$ thus consists of the column vectors of $A$ that are lined up together.

As proved in [16], if $C$ is a symmetrical matrix $(A^T A)$ formed from the measurements and $h$ is the parameter vector $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})^T$ given by the eigenvector of $C$ associated with the smallest eigenvalue, say $\lambda_1$, then the first-order perturbation in the parameter vector $h$ is given by

$$\delta_h \cong H \Delta H^T \Delta_{A^T A} h, \tag{5}$$

where

$$\Delta = diag\{0, (\lambda_1 - \lambda_2)^{-1}, (\lambda_1 - \lambda_3)^{-1}, (\lambda_1 - \lambda_4)^{-1}\}, \tag{6}$$

and $H$ is an orthonormal matrix, such that

$$H^{-1} C H = diag\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}. \tag{7}$$

$\Delta_{A^T A}$ is a $4 \times 4$ noise or perturbation matrix associated with $A^T A$. If $\Delta_{A^T A}$ can be estimated, then the perturbation $\delta_h$ in $h$ can be estimated by a first-order approximation as in (5).

We estimate $\Delta_{A^T A}$ from the perturbation in the surface measurements. We assume, for the sake of simplicity of analysis, that only the $z$ component of a surface measurement $X_i = (x_i, y_i, z_i)^T$ has errors, with this general model. This analysis is easily and directly extended to include errors in $x$ and $y$ if their noise variances are known.

Let $z_i$ have additive errors $\delta_{z_i}$ for $1 \le i \le n$. We then get

$$\Delta_A^T = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \delta_{z_1} & \delta_{z_2} & \cdots & \delta_{z_n} \\ 0 & 0 & \cdots & 0 \end{bmatrix}. \tag{8}$$

If the errors in $z$ at different points on the surface have the same variance $\sigma^2$, we get the covariance matrix

$$\Gamma_{A^T} = \sigma^2 diag\left(P_1, P_2, \cdots, P_n\right), \tag{9}$$

where $P_i$, $1 \le i \le n$, is a $4 \times 4$ submatrix:

$$P_i = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (10)$$

Now, consider the error in $h$. As stated before, we have

$$\delta_h \cong H\Delta H^T \Delta_{A^TA} h = H\Delta H^T \left[\mathcal{A}I_4 \ \mathcal{B}I_4 \ \mathcal{C}I_4 \ \mathcal{D}I_4\right]\delta_{A^TA} \triangleq G_h \delta_{A^TA}. \quad (11)$$

In the above equation, we have rewritten the matrix $\Delta_{A^TA}$ as a vector $\delta_{A^TA}$ and moved the perturbation to the extreme right of the expression. Then the perturbation of the eigenvector is the linear transformation (by matrix $G_h$) of the perturbation vector $\delta_{A^TA}$. Since we have $\Gamma_{A^T}$ $(= \Gamma_A^T)$, we need to relate $\delta_{A^TA}$ to $\delta_{A^T}$. Using a first-order approximation [16], we get

$$\Delta_{A^TA} \cong A^T \Delta_A + \Delta_A^T A. \quad (12)$$

Letting $A^T = [a_{ij}]^T \triangleq [A_1 \ A_2 \ \cdots \ A_n]$, we write

$$\delta_{A^TA} \cong G_{A^TA} \delta_{A^T}, \quad (13)$$

where $G_{A^TA}$ is easily determined from the equation $G_{A^TA} = [F_{ij}] + [G_{ij}]$, where $[F_{ij}]$ and $[G_{ij}]$ are matrices with $4 \times n$ submatrices $F_{ij}$ and $G_{ij}$, respectively; $F_{ij} = a_{ji} I_4$, and $G_{ij}$ is a $4 \times 4$ matrix with the $i$th column being the column vector $A_j$ and all other columns being zero. Thus, we get

$$\delta_h \cong G_h \delta_{A^TA} \cong G_h G_{A^TA} \delta_{A^T} \triangleq D_h \delta_{A^T}. \quad (14)$$

Then the covariance matrix of $h$ is given by

$$\Gamma_h \cong D_h \Gamma_{A^T} D_h^T. \quad (15)$$

The distance $d_s$ is affected by the errors in the estimation of the plane parameters, $h = (\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})^T$ and the $z$ measurement in $(x_i, y_i, z_i)^T$. Therefore, the error variance in $d_s$ is

$$\sigma_{d_s}^2 = \left[\frac{\partial d_s}{\partial \mathcal{A}} \ \frac{\partial d_s}{\partial \mathcal{B}} \ \frac{\partial d_s}{\partial \mathcal{C}} \ \frac{\partial d_s}{\partial \mathcal{D}} \ \frac{\partial d_s}{\partial z}\right] \times \left[\Gamma_{hs}\right] \times \begin{bmatrix} \frac{\partial d_s}{\partial \mathcal{A}} \\ \frac{\partial d_s}{\partial \mathcal{B}} \\ \frac{\partial d_s}{\partial \mathcal{C}} \\ \frac{\partial d_s}{\partial \mathcal{D}} \\ \frac{\partial d_s}{\partial z} \end{bmatrix}. \quad (16)$$

The covariance matrix $\Gamma_{hz}$ is given by

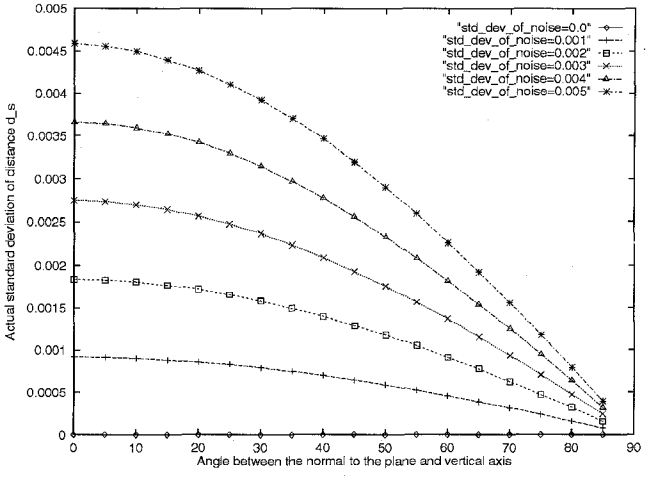
$$\Gamma_{hz} = \begin{bmatrix} \Gamma_h & 0 \\ 0 & \sigma^2 \end{bmatrix}. \quad (17)$$

Once the variance of $d_s$, $\sigma_{d_s}^2$ is estimated, we employ it in our optimal error function:

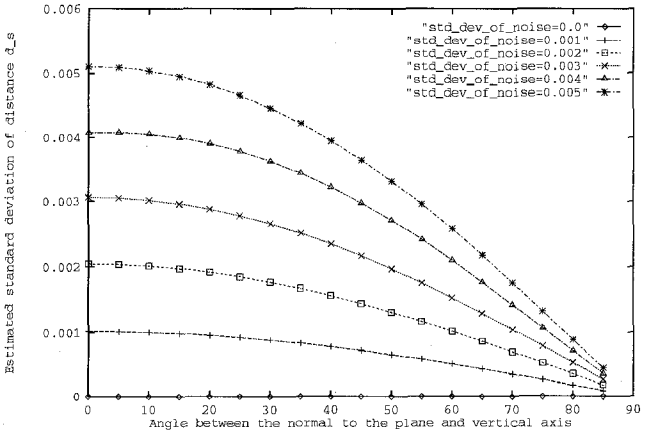
$$e^k = \sum_{i=1}^{N} \frac{1}{\sigma_{d_s}^2} d_s^2\left(\mathcal{T}^k \mathbf{p}_i, S_i^k\right). \quad (18)$$

#### 4.3.2 Simulation Results

Fig. 4a shows the plot of the *actual* standard deviation of the distance $d_s$ versus the orientation of the plane with respect to the vertical axis. Note that the mean of $d_s$ is zero when the surface points are in complete registration and when there is no noise. We generated two views of synthetic planar surfaces, with the view transformation between them being an identity transformation. We experimented with the planar patches at various orientations. We added uncorrelated Gaussian noise independently to the two views. Then, we estimated the distance $d_s$ at different control points using (2) and computed its standard deviation. The plot shows the values averaged over 1,000 trials. As indicated by our hypothesis, the actual standard deviation of $d_s$ decreases as the planar orientation goes from horizontal to vertical. As the variance of the added Gaussian noise to the $z$ measurements increases, $\sigma_{d_s}$ also increases. Similar results were obtained when we added uniform noise to the data [14].




(a)




(b)

Fig. 4. Standard deviation of $d_s$ versus the planar orientation with a Gaussian noise model: (a) actual $\sigma_{d_s}$; (b) estimated $\sigma_{d_s}$ using the perturbation analysis.

We compared the actual variance with the *estimated* variance of the distance ((16)) in order to verify whether our modeling of errors in $z$ values at various surface orientations is correct. We computed the estimated variance of the distance $d_s$ using our error model using (16) with the same experimental setup as described above. Fig. 4b illustrates the behavior of the estimated standard deviation of $d_s$ as the inclination of the plane (the surface orientation) changes. A comparison of Figs. 4a and 4b shows that both the actual and the estimated standard deviation plots have similar behavior with varying planar orientation, and their values are proportional to the amount of noise added. This proves the correctness of

our error model of $z$ and its effect on the distance $d_s$. Simulation results, when we repeated the experiments to compute both the *actual* and the *estimated* $\sigma_{d_s}$ using the planar parameters estimated with the linear regression method, were similar to those shown in Fig. 4. This also demonstrates the important fact that the method used for planar fitting does not bias our results.

## 5 View Registration Experiments

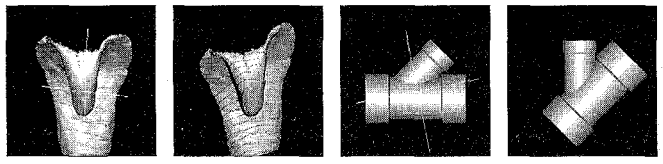
In this section, we demonstrate the improvements in the estimation of view transformation parameters on real range images using our MVE. We will henceforth refer to Chen and Medioni's technique [1] as the C-M method. We obtained range images of complex objects using a Technical Arts laser range scanner. We performed uniform subsampling of the depth data to locate the control points in view 1 that were to be used in the registration. From these subsampled points, we chose a fixed number of points that were present on smooth surface patches. The local smoothness of the surface was verified using the value of residual standard deviation resulting from the least-squares fitting of a plane in the neighborhood of a point. A good initial guess for the view transformation was determined automatically when the range images contained the entire object surface and the rotations of the object in the views were primarily in the plane. Our method is based on estimating an approximate rotation and translation by aligning the major (principal) axes of the object views [14]. Figs. 5a and 5c depict the two major axes of the objects. We used this estimated transformation as an initial guess for the iterative procedure in our experiments, so that no prior knowledge of the sensor placement was needed. Experimental results show the effectiveness of our method in refining such rough estimates. The same initial guess was used with the C-M method and the proposed MVE. We employed Newton's method for minimizing the error function iteratively.

In order to measure the error in the estimated rotation parameters, we utilize an error measure that does not depend on the actual rotation parameters. The relative error of rotation matrix $R$, $E_R$ is defined [16] to be $E_R = \left\|\tilde{R} - R\right\| \Big/ \|R\|$, where $\tilde{R}$ is an estimate of $R$. Since $RI = R$, the geometric sense of $E_R$ is the square root of the mean squared distance between the three unit vectors of the rotated orthonormal frames. Since the frames are orthonormal, $E_R = \sqrt{(dx^2 + dy^2 + dz^2)} \Big/ \sqrt{3}$. The error in translation, $E_t$, is defined as the square root of the sum of the squared differences between the estimated and actual $t_x$, $t_y$, and $t_z$ values.

### 5.1 Results

Fig. 5 shows the range data of a cobra head and a Big-Y pipe. The figure renders depth as pseudo intensity, and points almost vertically oriented are shown darker. View 2 of the cobra head was obtained by rotating the surface by 5° about the X axis and 10° about the Z axis. Table 1 shows the values of $E_R$ and $E_t$ for the cobra views estimated, using only as few as 25 control points. It can be seen that the transformation parameters obtained with the MVE are closer to the ground truth than those estimated using the unweighted objective function of the C-M method. Even when more control points (about 156) were used, the estimates using our method were closer to the ground truth than those obtained with the C-M method [14].

We also show the performance of our method when the two viewpoints are substantially different and the depth values are very noisy. Fig. 5 shows two views of the Big-Y pipe generated from the CAD model. The second view was generated by rotating the object about the Z axis by 45°. We also added Gaussian noise with mean zero and standard deviation of 0.5 mm to the $z$ values of the surfaces in view 2. Table 2 shows $E_R$ and $E_t$ computed with



(a) (b) (c) (d)

Fig. 5. Range images and their principal axes: (a) View 1 of a cobra head; (b) View 2 of the cobra head; (c) Big-Y pipe data generated from its CAD model; (d) View 2 of Big-Y pipe.

TABLE 1
Estimated Transformation for the Cobra Views

| Parameters | Actual value | C-M method | MVE |
|---|---|---|---|
| $\alpha$ (degrees) | 5 | 0.2857 | 4.5241 |
| $\beta$ (degrees) | 0 | –2.225 | 0.0143 |
| $\gamma$ (degrees) | 10 | 12.9741 | 10.3811 |
| $t_x$ (inches) | 0 | 0.9429 | 0.6684 |
| $t_y$ (inches) | 0 | 0.3483 | 0.0217 |
| $t_z$ (inches) | 0 | 0.2263 | –0.2976 |
| $E_R$ | | 0.0848 | 0.0087 |
| $E_t$ | | 1.0303 | 0.7319 |

154 control points. It can be seen from these tables that the transformation matrix, especially the rotation matrix, obtained with the MVE is closer to the ground truth than that obtained using the C-M method. The errors in translation components of the final transformation estimates are mainly due to the approximate initial guess. Our method refined these initial values to provide a final solution very close to the actual values. Our method also handled large transformations between views robustly. With experiments on range images of facial masks, we found that even when the depth data were quite noisy owing to the roughness of the surface texture of the objects and also due to self-occlusion, more accurate estimates of the transformation were obtained with the MVE. When the overlapping object surface between the views is quite small, the number of control points available for registration tends to be small, and, also, in such situations, the MVE has been found to have substantial improvement in the accuracy of the transformation estimate. Note that measurement errors are random, and we minimize the expected error in the estimated solution. However, our method does not guarantee that every component in the solution will have a smaller error in every single case.

TABLE 2
Estimated Transformation for the Big-Y Views

| Parameters | Actual value | C-M method | MVE |
|---|---|---|---|
| $\alpha$ (degrees) | 0 | 1.2636 | 0.6061 |
| $\beta$ (degrees) | 0 | 2.0151 | 1.1914 |
| $\gamma$ (degrees) | 45 | 44.4674 | 44.7786 |
| $t_x$ (inches) | 0 | 0.0163 | 0.0006 |
| $t_y$ (inches) | 0 | 0.1314 | 0.0845 |
| $t_z$ (inches) | 0 | 0.0438 | 0.0376 |
| $E_R$ | | 0.0348 | 0.0193 |
| $E_t$ | | 0.1395 | 0.0925 |

We also used the MVE for refining the pose estimated using cosmos-based recognition system for free-form objects [14]. The rotational component of the transformation of a test view of Vase2 (View 1 shown in Fig. 6a) relative to its best-matched model view (View 2 shown in Fig. 6b) was estimated using surface normals of

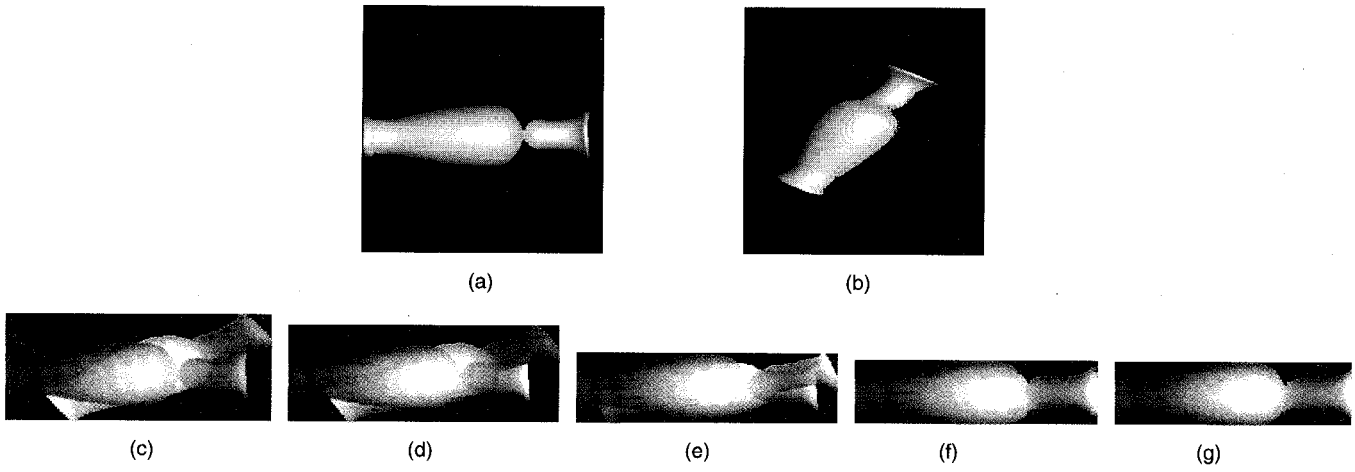



Fig. 6. Pose estimation: (a) View1 of a vase; (b) view2; (c)-(f) model view registered with the test view of Vase2 at the end of the first, third, fourth, and fifth iterations; (g) registered views at the convergence of the algorithm.

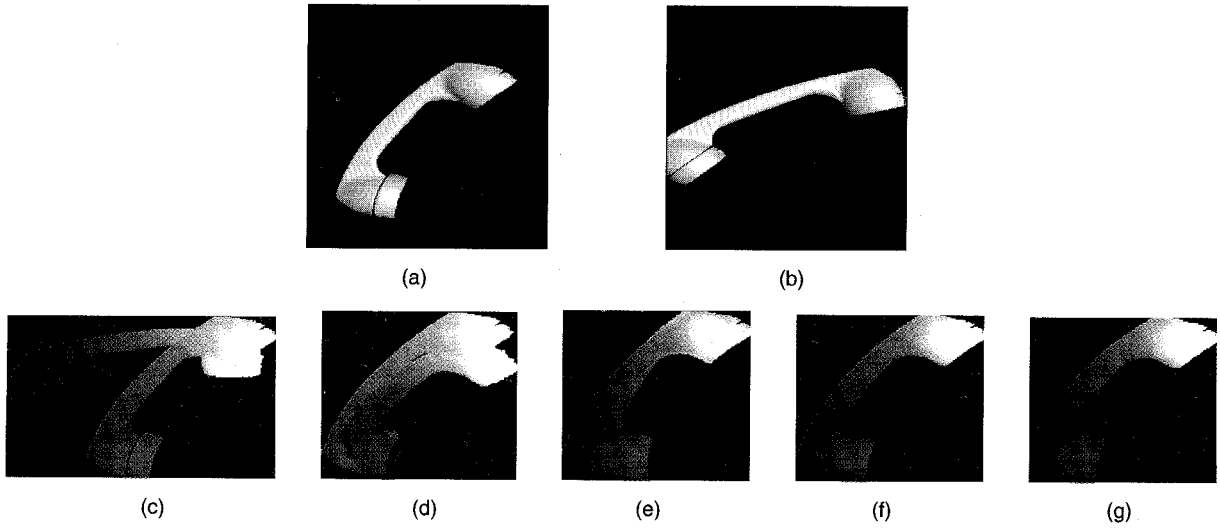



Fig. 7. Registration of views of Phone: (a) View 1; (b) view 2; (c)-(f) model view registered with the test view of Phone at the end of first, second, third, and fourth iterations; (g) registered views at the convergence of the algorithm.

*corresponding* surface patch-groups determined by the recognition system. A total of 10 pairs of corresponding surface patch-groups was used to estimate the average rotation axis and the angle of rotation. These rotation parameters ($\mathbf{r}$ = (0.005288, −0.004433, 0.024552) and $\theta = 0.180429$ radians) were used to compute the $3 \times 3$ rotation matrix, which was then used as an initial guess to register the model view (View 2) with the test view (View 1) of Vase2 using the MVE. We note here that the computational procedure for MVE was augmented using a verification mechanism for checking the validity of the control points during its implementation [17]. We derived the results presented in this section using this augmented procedure. Figs. 6c, 6d, 6e, and 6g show the iterative registration of the model view with the scene view. It can be seen that the views are in complete registration with one another at the end of seven iterations.

Fig. 7 shows the registration of a model view with a scene view of Phone through several iterations of the algorithm. The registration scheme converged with the lowest error value at the sixth iteration. It can be seen that, even with a coarse initial estimate of the rotation, the registration technique can align the two views successfully within a few iterations. Given a coarse correct initial guess, registration, on the average, takes about 30 seconds to register two range images whose sizes are 640 × 480 on a SPARCstation 10 with 32MB RAM.

### 5.2 Discussion

In general, all the orientation parameters of an object will be improved by the proposed MVE method if the object surface covers a wide variety of orientations, which is true with many natural objects. This is because each locally flat surface patch constrains the global orientation estimate of the object via its surface normal direction. For example, if the object is a flat surface, then only the global orientation component that corresponds to the surface normal can be improved, but not the other two components that are orthogonal to it. For the same reason, the surface normal of a cylindrical surface (without end surfaces) covers only a great circle of the Gaussian sphere, and, thus, only two components of its global orientation can be improved. The more surface orientations that an object covers, the more complete the improvement in its global orientation can be, by the proposed MVE method. An analysis of the performance of the MVE and unweighted registration algorithms with surfaces of various geometries can be found in [14].

When more than two views have to be registered, our algorithm for registering a pair of object views can be used either sequentially (with the risk of error accumulation) or in parallel, e.g., with the star-network scheme [9]. Note, however, that we have not extended our weighted approach to the problem of computing the transformation between $n$ views simultaneously. When there is a

significant change in the object depth, the errors in z at different points on the surface may no longer have the same variance; the variance typically increases with greater depth. In such situations, our perturbation analysis still holds, except for the covariance matrix $\Gamma_{A^T}$ in (9). The diagonal elements of this matrix will no longer be identical, as we assumed. Each element, which is a summary of the noise variance at the corresponding point in the image, must reflect the combined effect of variation due to depth, measurement unreliability due to surface inclination, etc., and, therefore, a suitable noise model must be assumed or experimentally created.

## 6 Summary

Noise in surface data is a serious problem in registering object views. The transformation that relates two views should be estimated robustly in the presence of errors in surface measurements for seamless view integration. We established the dependency between the surface orientation and the accuracy of surface normal estimation in the presence of error in range data, and its effect on the estimation of transformation parameters with geometrical analysis and experimental results. We proposed a new error model to handle uncertainties in z measurements at different orientations of the surface being registered. We presented a first-order perturbation analysis of the estimation of planar parameters from surface data. We derived the variance of the point-to-plane distance to be minimized to update the view transformation during registration. We employed this variance as a measure of the uncertainty in the distance resulting from noise in the z value, and proposed a minimum variance estimator to estimate transformation parameters reliably. The results of our experiments on real range images have shown that the estimates obtained using our MVE generally are significantly more reliable than those computed with an unweighted distance criterion.

## Acknowledgments

This work was supported by a grant from Northrop Corporation. We thank the reviewers for their helpful suggestions for improvement.

## References

[1] Y. Chen and G. Medioni, "Object Modelling by Registration of Multiple Range Images," *Image and Vision Computing*, vol. 10, pp. 145-155, Apr. 1992.

[2] F.P. Ferrie and M.D. Levine, "Integrating Information from Multiple Views," *Proc. IEEE Workshop on Computer Vision*, pp. 117-122, Miami Beach, Fla., 1987.

[3] M. Potmesil, "Generating Models for Solid Objects by Matching 3D Surface Segments," *Proc. Int'l Joint Conf. Artificial Intelligence*, pp. 1,089-1,093, Karlsruhe, Germany, 1983.

[4] P.J. Besl and N.D. McKay, "A Method for Registration of 3-D Shapes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239-256, Feb. 1992.

[5] G. Turk and M. Levoy, "Zippered Polygon Meshes from Range Images," *Proc. SIGGRAPH '94*, pp. 311-318, 1994.

[6] Z. Zhang, "Iterative Point Matching for Registration of Free-Form Curves and Surfaces," *Int'l J. Computer Vision*, vol. 13, no. 2, pp. 119-152, 1994.

[7] G. Blais and M.D. Levine, "Registering Multiview Range Data to Create 3D Computer Graphics," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 820-824, Aug. 1995.

[8] R. Bergevin, D. Laurendeau, and D. Poussart, "Registering Range Views of Multipart Objects," *Computer Vision and Image Understanding*, vol. 61, pp. 1-16, Jan. 1995.

[9] R. Bergevin, M. Soucy, H. Gagnon, and D. Laurendeau, "Towards a General Multi-View Registration Technique," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 5, pp. 540-547, May 1996.

[10] R.M. Haralick, H. Joo, C.-N. Lee, X. Zhuang, V.G. Vaidya, and M.B. Kim, "Pose Estimation from Corresponding Point Data," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 19, no. 6, pp. 1,426-1,446, 1989.

[11] P. Hebert, D. Laurendeau, and D. Pousart, "Scene Reconstruction and Description: Geometric Primitive Extraction from Multiple View Scattered Data," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 286-292, New York, 1993.

[12] P.J. Flynn and A.K. Jain, "Surface Classification: Hypothesis Testing and Parameter Estimation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 261-267, Ann Arbor, Mich., 1988.

[13] T.S. Newman, "Experiments in 3D CAD-Based Inpection Using Range Images," PhD thesis, Michigan State Univ., Dept. of Computer Science, 1993.

[14] C. Dorai, "COSMOS: A Framework for Representation and Recognition of 3D Free-Form Objects," PhD thesis, Dept. of Computer Science, Michigan State Univ., May 1996.

[15] J. Weng, P. Cohen, and N. Rebibo, "Motion and Structure Estimation from Stereo Image Sequences," *IEEE Trans. Robotics and Automation*, vol. 8, pp. 362-382, June 1992.

[16] J. Weng, T.S. Huang, and N. Ahuja, "Motion and Structure from Two Perspective Views: Algorithms, Error Analysis, and Error Estimation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 5, pp. 451-476, May 1989.

[17] C. Dorai, G. Wang, A.K. Jain, and C. Mercer, "From Images to Models: Automatic 3D Object Model Construction from Multiple Views," *Proc. 13th Int'l Conf. Pattern Recognition*, vol. I, pp. 770-774, Vienna, Aug. 1996, submitted to *IEEE Trans. Pattern Analysis and Machine Intelligence*.