Estimating pose uncertainty for surface registration

A. J. Stoddart, S. Lemke, A. Hilton, T. Renn
Dept. of Electronic & Electrical Engineering
University of Surrey, Guildford, Surrey GU2 5XH, UK
email: a.stoddart@ee.surrey.ac.uk

Abstract

Accurate registration of surfaces is a common problem in computer vision. Several algorithms exist to refine an approximate value for the pose to an accurate value. They are all more or less variants of the Iterated Closest Point algorithm of Besl and McKay (1992).

Up to now the problem of determining the uncertainty in the pose estimate thus obtained has not been addressed in detail. In this paper we present a framework in which to quantify the uncertainty in pose. We introduce a new parameter called the *registration index* to give a simple means of quantifying the pose errors one might expect when registering a particular shape.

1   Introduction

The registration of surfaces and curves can be broken down into two separate tasks. A *matching* task in which an approximate estimate of the pose is obtained. This might also involve a large database of model shapes. Then a *registration* task in which an accurate pose is obtained from a rough initial guess.

Algorithms for the accurate registration task are now well established. They are all more or less variations on the iterative closest point algorithm (ICP). The algorithm is not particularly fast or simple, but can give very accurate results. It is difficult to imagine a substantially different or better algorithm being developed for the general case of free form surfaces.

Although one may obtain a very accurate estimate of pose, very little work exists to *quantify* the accuracy. Recently there has been some work by Pennec and Thirion (1995a, 1996) which has provided a very useful framework for quantifying uncertainty in pose.

In this paper we apply this framework to the problem of surface registration. The original contribution of this paper is to propose a framework for quantifying the uncertainty in pose when two surfaces have been registered. We also introduce a simple parameter which depends only on shape which we call the *registration index*. This gives an easy way to predict the expected quality of the pose estimate under conditions of varying number of point measurements and sensor noise.

The most closely related work to this is that of Brujic and Ristic who performed a Monte Carlo study of extracted pose parameters (1996).

1.1 Besl's ICP Algorithm

The ICP algorithm was first presented by Besl and Mackay (1992). The algorithm may be summarised as follows. Using a reasonably good initial guess the scene and model are moved into approximate alignment. A set of points is chosen on one surface, and the corresponding closest points are found on the other surface. Using a least squares method the pose is computed that brings the two point sets into alignment, and this is applied to one of the surfaces. The procedure is iterated until the change in pose becomes very small.

It should be noted that the algorithm will find a nearest local minimum of a mean-square distance metric. It is mainly in this respect that the initial estimate is important, since we would prefer to find the global minimum.

The usually quoted advantages of the ICP method are as follows. It handles the full six degrees of freedom problem. It is independent of shape representation (assuming that you have a closest point algorithm). It does not require derivative estimation or feature extraction. The main disadvantage is the requirement for a good initial guess.

Two distinct problems may be addressed by "surface registration" algorithms. The model-scene case arises when we have an exact model and wish to register it to scene measurements. The scene-scene case arises when we wish to register (and possibly fuse (Hilton et al., 1996)) two separate scene measurements of the same object. The registration method is usually the same but for a correct statistical analysis of the errors the two cases need to be treated slightly differently.

We should note that most variants of the ICP that have been presented have used the least squared distance point set registration procedure. This is probably due to the ready availability of an analytical solution. However Chen and Medioni (1992) use a registration procedure that includes reference to local surface normals. This is not analytically soluble but seems better adapted to the problem.

Later work by Dorai *et al* (1994) revisited the technique of Chen and Medioni, and proposed an minimum variance estimator. This accepted the approach of using only uncertainty in the surface normal direction, but argued that range data has constant uncertainty in the z direction, which implies slope dependent perpendicular uncertainty. This can be accommodated within our scheme.

Applications of registration should be mentioned. One area is that of CAD based inspection (Bispo and Fisher, 1994; Brujic and Ristic, 1996). It is assumed that there is an accurate model, perhaps based on B-Splines of an object, and a surface measurement by a range sensor. The objective is to register the data to the model and then find the largest or average deviation of the object from the model. Other applications include navigation and surface fusion (Hilton et al., 1996).

2 Notation

In order to present our notation we begin by considering the familiar problem of pose estimation from two point sets with correspondence. Various analytical solutions exist, and we use the SVD method of (Kanatani, 1994).

A pose consists of a translation t and a rotation. The rotation may be specified

by an axis $\hat{n}$ and angle θ. Equivalently it may be specified by the 3×3 rotation matrix R or the 3 component vector $r = \theta\hat{n}$. Finally there is the quaternion representation q. In this paper we will mainly use r, thus a pose or rigid body transform f is equivalent to

$$f = \left| \begin{array}{c} r \\ t \end{array} \right. = (R, t) \tag{1}$$

Poses are also referred to as frames. Frames may be composed $\circ$ or applied $*$ to points.

- Application of $f = (R, t)$ to x : $y = f * x = Rx + t$.

- Composition of $f_1 = (R_1, t_1)$ with $f_2 = (R_2, t_2)$:
  $f = f_2 \circ f_1 = (R_2 R_1, R_2 t_1 + t_2)$

Suppose we have two sets of N corresponding points b_i and q_i. A pair (b_i, q_i) together comprise a measurement $\hat{x}_i$ of some underlying unknown true value x_i. We will assume an unbiased measurement so that $\overline{x} = E[\hat{x}] = x$ and denote the covariance by $W_x = E[(\hat{x} - \overline{x})(\hat{x} - \overline{x})^\top]$.

We assume that the true values of b_i and q_i are related by an unknown frame f according to the relation $q_i = f * b_i$. We need to define an error vector h_i with covariance W_i such that the estimate $\hat{f}$ that we seek will minimise the weighted least squares criterion

$$C = \sum_i h_i^\top W_i^{-1} h_i \tag{2}$$

A suitable error measure is the vector distance $h_i(x_i, f) = q_i - f * b_i$, so that we minimise the mean squared distance. This is a non-linear minimisation but exact solutions are available. For greater generality we choose to linearise the minimisation problem in the variable f. It can then be solved either directly or using the Extended Kalman Filter (EKF). These methods can output an estimate of the transform $\hat{f}$ and the covariance of the transform W_f. The criterion is linearised around some prior estimate $\hat{f}_{k-1}$ as follows

$$h_i(x_i, f) \simeq h(\hat{x}_i, \hat{f}_{k-1}) + \frac{\partial h_i}{\partial f}(f - \hat{f}_{k-1}) + \frac{\partial h_i}{\partial x_i}(x_i - \hat{x}_i) \tag{3}$$

If we identify

$$Y_i = -h_i(x_i, f) + \frac{\partial h_i}{\partial f} f \tag{4}$$

$$M_i = \frac{\partial h_i}{\partial f} \tag{5}$$

$$w_i = \frac{\partial h_i}{\partial x_i}(x_i - \hat{x}_i) \tag{6}$$

Then we have a standard multidimensional least squares problem of the form

$$Y_i = M_i f + w_i \tag{7}$$

Y_i is the generalised measurement, f is the parameter vector and w_i the noise, with covariance W_i. In this paper we assume that the point measurement noise is always isotropic, i.e. $W_b = \rho^2 I_3$. The points b_i are always measurements or scene points. The q_i may be (noise free) model points or scene points. In the scene-scene case $W_i = 2\rho^2 I_3$ and in the model-scene case $W_i = \rho^2 I_3$. We will sometimes refer to $\sigma^2 = 3\rho^2$, where σ is the rms displacement counted over all 3 dimensions.

The minimum of

$$C = \sum_{i=1}^{N}(Y_i - M_i f)^{\top} W_i^{-1}(Y_i - M_i f) \tag{8}$$

is given by

$$\hat{f}_k = \left[\sum_{i=1}^{N} M_i^{\top} W_i^{-1} M_i\right]^{-1} \left[\sum_{i=1}^{N} M_i^{\top} W_i^{-1} Y_i\right] \tag{9}$$

The covariance of the estimated parameter $\hat{f}_k$ will be given by

$$W_f^{-1} = \sum_{i=1}^{N} M_i^{\top} W_i^{-1} M_i \tag{10}$$

The result depends on $\hat{f}_{k-1}$ because this is the point around which the cost was linearised. Therefore we must iterate a few times until the estimate $\hat{f}_k$ converges. The Extended Kalman filter provides a good framework to solve this problem . Finally we need to evaluate the partial derivatives. This is not straightforward due to the nonlinear nature of rotations, and for details we refer the reader to (Pennec and Thirion, 1995b).

2.1 Frame covariance

This resolves the question of how to compute frame covariance. However we must ask whether it makes sense. Pennec and Thirion (1995b) suggest that the correct way to model frame errors is by a right error frame e_r with some probability distribution $P(e_r)$. This is a representation independent concept since the e_r are real rigid body transforms. The frame estimate is then given by $\hat{f} = f \circ e_r$. This is better than an additive noise model such as $\hat{f} = f + \delta f$ where it is difficult to ensure that the noise process does the same thing to different frames f. In addition it will be specific to a particular representation of f. These problems are related to the fact that rotations do not belong to a vector space. For further discussion see (Pennec and Thirion, 1995b).

In the applications that we will consider the errors are typically small so that the rotation errors are much less than $90°$. Infinitesimal rotations do form a vector space. This means that for very small frames

$$\delta f_1 \circ \delta f_2 \simeq \delta f_1 + \delta f_2 \tag{11}$$

in the representation of equation (1), namely $f = (r, t)$. It does *not* apply in other representations. Thus if we move the scene and model to a position where both f

and the typical error are small then

$$W_f = E[(\hat{f} - \overline{f})(\hat{f} - \overline{f})^\top] \simeq E[e_r e_r^\top] \tag{12}$$

Henceforth in this paper we will assume that this has been done.

2.2 Interpretation of the frame covariance

The covariance W_f consists of 36 real numbers and is not immediately useful as a way of grasping the uncertainty in the pose. Covariance matrices may be diagonalized to reveal a set of directions in which the errors are uncorrelated, with their corresponding variances.

The infinitesimal frame error (usually) consists of 3 small rotations and 3 small translations. If a specific degree of freedom is very uncertain the covariance will be large. After diagonalisation we can identify the largest error modes, for example a pure translation in a particular direction, a pure rotation or a mixture.

The following example will illustrate the interpretation of the covariance matrix. In this example we consider only the rotational modes. A cylinder of length l and radius r is centred at the origin, and aligned with direction $\hat{a}$. It is filled with a uniform distribution of N points. We create a copy and corrupt each point of the copy by Gaussian noise with variance $\rho^2 I_3$ and then obtain an estimate of (f, W_f). If $r \ll l$ we would expect the frame to be accurately determined for rotations perpendicular to the axis and poor for the rotations around the axis.

This is related to the well known fact in point set registration that some solutions are degenerate. If one set of points lie in a straight line the solution is defined up to a rotation around the line. The analytical methods presented by Kanatani (1994) are able to detect the presence of degenerate solutions, but cannot detect cases of near degeneracy such as points contained in a narrow cylinder.

If we diagonalise the 3×3 rotational submatrix then we get eigenvalues as follows

$$\lambda_{1,2} \simeq \frac{\rho^2}{N} \frac{12}{l^2}, \qquad \lambda_3 \simeq \frac{\rho^2}{N} \frac{2}{r^2} \tag{13}$$

This is in agreement with the idea that there are two low covariance modes of rotation perpendicular to the axis and a high covariance (uncertain) mode of rotation parallel to the axis. Indeed the third eigenvector is the same as $\hat{a}$. In the limit of $r \to 0$ the covariance blows up consistent with a degenerate solution.

3 Surface Registration

We wish to apply the ideas of (Pennec and Thirion, 1995b) to the case of surface registration. At the heart of most ICP algorithms is the use of the least squares criterion of equation (2). There is a serious problem with this which is best illustrated by example. Suppose we have as our model a sphere and have several accurate measurements from its surface. Using the ICP approach for each measurement we find the closest point on the model and then obtain the registration from the two point sets with correspondence. The covariance produced by the simple criterion will tell us that we have a very accurate estimate of pose.

Intuitively we know that we can determine only 3 of the 6 degrees of freedom of the pose, the remaining 3 are degenerate. We can see that the complete surface criterion will be unaffected by rotations of the model around its centre.

This strongly suggests that we should use a different error measure which takes account of the origin of the model-data point pairs. In particular the way in which we obtain correspondence does not guarantee "true" correspondence.

Another approach is to consider an idealised surface sensor. Suppose we have a surface consisting of the x-y plane and we measure a point $(x, y, 0)$ on it which is then corrupted by some small error to become $(x + \delta x, y + \delta y, \delta z)$. An ICP type algorithm will regard this point as a measurement of $(x + \delta x, y + \delta y, 0)$, corrupted by an error $(0, 0, \delta z)$, since we have no way of knowing which point of the surface to associate with it. Only the z component of the measurement has any bearing on the pose estimation.

This suggests that we use only the component of the point-to-surface distance in the direction of the model surface normal when defining the least squares error criterion. This difference may affect the minimum of the cost function considerably. Most other authors, with the exception of Chen and Medioni (1992), do not use the surface normal. Chen and Medioni use a motivation roughly similar to our own, although their method differs in detail.

Notice here that we assume that the sensor measures points and that the model is a complete surface description. Of course in the scene-scene case the surface would be reconstructed from the point measurements and the normal would be noisy. This noise will be taken into account in future work.

3.1 Point Set plus normal Registration

The discussion indicates we should use an error model for the inner registration step of the ICP which uses the surface normal. We assume that we have points m_i on the model which are translated by frame f^{-1} and corrupted by noise to become data points d_i, i.e. $d_i = f^{-1} * m_i + w_i$. We assume that the noise is isotropic so that $m_i = f * d_i + w_i$.

We assume that the model surface normal at point m_i is n_i and noise free. The error measure should only consider errors perpendicular to the surface so we propose $h'_i = n_i.(m_i - f * d_i)$ with the error criterion given by

$$C = \sum_i h_i'^\top W_i^{-1} h_i' \tag{14}$$

The analysis of section 2 may now be applied to the new error function. The differences are relatively minor, except that we should replace M_i by $M'_i = n_i.M_i$. We now consider an example. Suppose that the model is a unit square centred on the origin with normal $n = (0, 0, 1)$. We select N points at random, apply some f and corrupt with Gaussian noise of variance σ^2. As expected the inverse covariance matrix has zeros in the rows and columns corresponding to rotations around the $\hat{z}$ axis and translations in the x-y plane. The covariance blows up indicating that the recovered frame is 3-fold degenerate. A similar exercise may be performed with a sphere. In this case the translations are well fixed but there are three undetermined rotations.

3.2   Registration Index

In many situations we will not be interested in analysing in detail the covariance matrix. In this section we propose a simple parameter, the registration index r_s, that will give an indication of how well two surfaces may be registered. In particular it will signal cases of near degeneracy. The definition of r_s is based on the *typical boundary error* σ_8 as proposed by Pennec and Thirion (1995a). This is a concise way of expressing the pose uncertainty.

Typically we are not interested in the pose uncertainty but rather its effects on point registration. The error in a point's registration is best expressed by the rms distance. Suppose that we have $\hat{y} = \hat{f} * x$, i.e. x known and $\hat{f}$ an estimate. The variance in estimate $\hat{y}$ due to pose uncertainty W_f is

$$W_y = J W_f J^\top, \qquad J = \frac{\partial (f * x)}{\partial f} = -M_i \qquad (15)$$

The rms distance of y from $\hat{y}$ due to the uncertainty in f we denote by σ_d where

$$\sigma_d^2 = E[(\hat{y} - y)^2] = trace W_y \qquad (16)$$

In this way we see that the error of a point depends on the pose uncertainty. However the point error also depends on where the point is. A convenient way of reducing the dependence on position is to use the 8 vertices of a bounding box. This will usually give a good indication of the average error of points inside the box. Thus we define the typical boundary error σ_8 as the arithmetic average of the eight rms displacements (i.e. σ_d) of points at the vertices of the bounding box. For fixed W_f it depends on the size of the box.

We have not yet said where W_f comes from. Consider the following scenario: We have a cube centred at the origin with sides of length l. We obtain $N/8$ measurements of each vertex with $W = \rho^2 I_3$. We then use point set registration to compute $(\hat{f}, W_f(cube))$, and obtain $\sigma_8^2(cube) = K_c \rho^2 / N$ where K_c is a number independent of the cube size, N and ρ. It may be determined numerically that $K_c = 6$. For any positioning of the measured points closer to the cube centre the rotational errors and hence K_c will increase. $K_c = 6$ is the smallest (i.e. optimal) value if the pose is computed from measurements of points located within the cube.

We can now select some surface registration problem and compute the pose using our modified ICP algorithm $(\hat{f}, W_f(surface))$. Based on this we can compute $\sigma_8^2(surface)$. (The box used to compute σ_8 should be a bounding box of the intersection of the surfaces for partially overlapping surfaces.) We now define K_s by the relation $\sigma_8^2(surface) = K_s \rho^2 / N$.

K_s will be larger than K_c and will blow up as the registration problem becomes degenerate. We define the registration index r_s as the ratio

$$r_s = \frac{K_s}{K_c} \qquad (17)$$

The larger the registration index the less certain the pose. Knowledge of the registration index allows a "back-of-the-envelope" computation of the typical boundary

Table 1: Registration Indices

Shape	r_s
Sphere	∞
Tetrahedron	6.0
Octahedron	19.0
Cube	6.0
Dodecahedron	20.0
Icosahedron	24.9
Beethoven	11.1
Foot	11.9
Dumbbell	20.9
Arch	12.4
Bunny	13.3

error, given N measurements and measurement error of ρ. The registration index converts to typical boundary error as follows

$$\sigma_8^2 = 6 r_s \frac{\rho^2}{N} \tag{18}$$

3.3 Surface Registration Examples

The first examples we consider are a sphere radius r and the five platonic solids inscribed within a sphere radius r. We consider first the icosahedron and select N points at random, apply some frame f and add noise with variance ρ^2. The ICP algorithm is applied to recover an estimate $\hat{f}$. If we use a cost based on h we obtain translational modes with covariance ρ^2/N whereas the rotational modes have covariance $2\rho^2/Nr^2$. If we instead use h' we obtain $3\rho^2/N$ and $30\rho^2/Nr^2$ respectively. This shows that the icosahedron does resemble the sphere to some extent and the rotational modes have relatively high covariance as a consequence. For the icosahedron we obtain $r_s \simeq 24.9$. The registration indices are summarized in Table 1. For the sphere $r_s = \infty$. In figures 1 and 2 we show the registration index for a variety of other shapes from various sources.

Our area of interest lies in surface fusion where the model is unknown and we have two scene measurements. In the case where we are registering a scene to a model we have access to the exact normal. If we are registering two measurements of the same object we do not. The normal is a differential quantity and therefore magnifies the effects of measurement error.

Work is in progress on developing the treatment further to take account of the error in the normal, but for the time being we believe that some indication of the error may be obtained from naively applying the model scene method.

In figure 3 we show a scene-scene registration task and record the registration indices. In table 1 we summarise all the registration indices. A 'fringe benefit' of this technique should be mentioned. We find that an ICP based on the point-
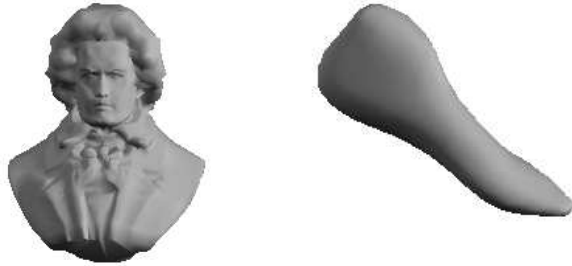
Figure 1: (a) Beethoven $r_s = 11.1$ (b) Foot $r_s = 11.9$
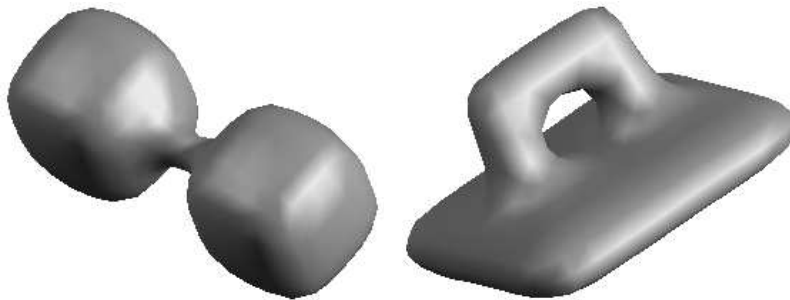
Figure 2: (a) Dumbbell $r_s = 20.9$ (b) Arch $r_s = 12.4$

normal registration usually needs between 1/10th and 1/2 the number of iterations. The reason for this may be understood by considering the following example. Consider a cube translated parallel to a side. The point-normal ICP will converge in one step, whereas the point-point ICP will converge in a geometric series where the rate depends on the number of points on faces perpendicular and parallel to the displacement. The points on parallel faces slow down the point-point ICP convergence, but are ignored in the point-normal ICP.

4 Conclusion

We have presented a method to compute the frame covariance for surface registration. We believe that it favours the Chen and Medioni variant of the Iterated Closest Point algorithm. We find that the ICP algorithm with this variant is 2-10 times faster than the conventional choice.

Work is in progress on the case of curve-curve registration where the error function should be modified to $h'' = t \times h$ for t the tangent vector. The treatment is otherwise the same. The treatment we have presented is complete for the case of model-scene registration where we have access to an exact surface normal. Work is in progress on the case of scene-scene registration where the uncertainty in the normal should be taken into account.

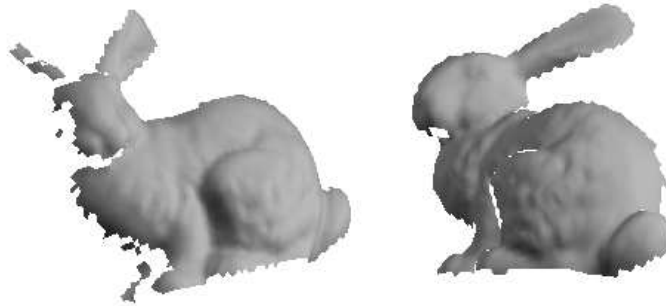Finally we would like to acknowledge the sources of the data. The foot model

Figure 3: The scene and model for the bunny, $r_s = 13.3$

is based on Cyberware data supplied by Tim McInerney, using the Slime package. The dumbbell and arch were also created using Slime. The Beethoven model were produced by Viewpoint Animation using a mechanical digitiser. The bunny dataset was used by Turk and Levoy in their zipper paper.

References

Besl, P.J. and McKay, N.D. (1992). A method for registration of 3-d shapes. *IEEE Trans. Pattern Analysis and Machine Intell., 14*, no. 2, 239–256.

Bispo, E. M. and Fisher, R. B. (1994). Free-form surface matching for surface inspection. In *6th IMA Conf. on the Mathematics of Surfaces*, pp. ?–?, Brunel University, London, U.K.

Brujic, D. and Ristic, M. (1996). Analysis of free form surface registration. In *1996 Int. Conference on Image Processing*, pp. ?–?, Laussane, Switzerland.

Chen, Y. and Medioni, G. (1992). Object modelling by registration of multiple range images. *Image and Vision Computing, 10*, no. 3, 145–155.

Dorai, C., Weng, J., and Jain, A.K. (1994). Optimal registration of multiple range views. In *12th Int. Conference on Pattern Recognition*, pp. A569–571, Jerusalem, Israel.

Hilton, A., Stoddart, A. J., Illingworth, J., and Windeatt, T. (1996). Reliable surface reconstruction from multiple range images. In *Fourth European Conference on Computer Vision*, pp. 117–126, Cambridge, U.K.

Kanatani, K. (1994). Analysis of 3-d rotation fitting. *IEEE Trans. Pattern Analysis and Machine Intell., 16*, no. 5, 543–549.

Pennec, X. and Thirion, J.-P. (1995a). Validation of 3-d registration methods based on points and frames. In *5th Int. Conference on Computer Vision*, pp. 557–562, Cambridge, Massachusetts.

Pennec, X. and Thirion, J.-P. (1995b). Validation of 3-d registration methods based on points and frames. INRIA Tech Report RR2470.

Pennec, X. and Thirion, J.-P. (1996). A framework for uncertainty and validation of 3-d registration methods based on points and frames. *Int. Journal of Computer Vision*.