Registration of Multiple Point Sets

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

Abstract

Registering 3D point sets subject to rigid body motion is a common problem in computer vision. The optimal transformation is usually specified to be the minimum of a weighted least squares cost. The case of 2 point sets has been solved by several authors using analytic methods such as SVD. In this paper we present a numerical method for solving the problem when there are more than 2 point sets.

Although of general applicability the new method is particularly aimed at the multiview surface registration problem. To date almost all authors have registered only two point sets at a time. This approach discards information and we show in quantitative terms the errors caused.

1. Introduction

Registering point sets subject to rigid body motion is a common problem in computer vision. It occurs in many applications where 3-D data are obtained by stereo, range sensors, tactile sensing, etc. It is also encountered in problems such as registering free-form curves or surfaces. The method presented in this paper is of immediate relevance to model building from multiple range images. Free-form surfaces taken from multiple viewpoints must be registered prior to fusion.

For two 3-D point sets with known correspondences the optimal alignment is usually defined as the minimum of an appropriate cost. Usually the cost is a weighted sum of the squared Euclidean distances between each pair of corresponding points.

In the case that there are 3 rigid objects or views with pairwise point correspondences between set 1 and 2, and set 1 and 3 the problem breaks down into 2 binary problems. However if, in addition, correspondences are provided between set 2 and set 3 the pairwise approach no longer produces the optimal result. If the data is noisy this may result in significant and unnecessary error.

In this paper we present a simple method to obtain the optimal solution to the multiple point set registration problem. The method is numerical and iterative.

2. Literature Review

Several analytic solutions to the problem of registering two point sets have been presented. For a discussion of the various methods see Kanatani [3] and references therein. In a recent paper [4] the various methods were compared. It was concluded that the various analytical methods give answers in good agreement with each other for non-degenerate data sets.

Point set registration is needed in many applications in range imaging. One application is CAD based inspection where a point set is registered to a CAD model. Another common problem is building complete surface models of free-form objects from multiple range images. Free-form surfaces taken from multiple viewpoints must be registered prior to fusion. Typically more than 10 views are needed to reconstruct an entire object and each view will overlap with several neighbours [5, 6].

Most range image registration algorithms are variations of the Iterative Closest Point (ICP) algorithm [1]. Most published algorithms study the case of registering 2 surfaces at a time. There are many situations where there are more than two point sets to be combined. Turk and Levoy [6] address a surface fusion problem where multiple surfaces may be combined. In the absence of a suitable technique for multiple point set fusion they register all views to a single master view. The master view is a cylindrical scan and the secondary views scanned over an x-y grid.

In this way they reduce the multiple fusion problem to several binary problems that can be solved by existing methods. This is however not altogether satisfying. The solution can be improved by considering also the overlap between pairs of secondary views. The improvement may be very slight, but range data is often used in applications where accuracy is important. The problem of multiple view registration has also been addressed by Gagnon *el al* [2].

Most surface registration algorithms [1, 7, 6] generalize immediately to multiple surfaces provided that a multiple point set registration algorithm is available. It is for applications such as this that we present our method.

3. Problem Specification

We assume that there are M distinct rigid bodies. Each rigid body could correspond to a view (range image) of an object. We will denote views by α or $\beta = 1..M$. The points originating from view α are denoted by $\hat{x}^\alpha$ a three component column vector. The $\hat{x}^\alpha$ may be transformed by a 3×3 rotation matrix R^α followed by a translation T^α into some point $\hat{y}^\alpha$. We write

$$\hat{y}^\alpha = R^\alpha \hat{x}^\alpha + T^\alpha \tag{1}$$

In this paper there is one fixed coordinate system and the transformations act on the point sets.

Next we suppose that there are P sets of pairwise correspondences denoted by $\mu = 1..P$. The μth set of pairwise correspondences is a set of correspondences between points from view $\alpha(\mu)$ and points from view $\beta(\mu)$. The number of points in pair set μ is denoted N_μ. The ith point pair in pair set μ consist of a point $\hat{x}_i^{\alpha(\mu)}$ from view $\alpha(\mu)$ and a point $\hat{x}_i^{\beta(\mu)}$ from view $\beta(\mu)$. The weighting assigned to this pair is denoted w_i^μ.

Different correspondence sets with points from the same view may or may not use the same points from that view. In other words there may be as few as M *distinct* point sets $\{\hat{x}_i^{\alpha(\mu)} | i = 1..N_\mu\}$ or as many as $2P$.

We assume that we are given all the points $\hat{x}_i^{\alpha(\mu)}$ and $\hat{x}_i^{\beta(\mu)}$ and the weights w_i^μ. We must now find the M optimal rotations and translations for the M views, where the quantity we wish to optimize will now be defined. The objective function $E[R^1..R^M, T^1..T^M]$ is a weighted sum of squared Euclidean distance defined as

$$E[R^1..R^M, T^1..T^M] = \sum_{\mu=1}^{P} \sum_{i=1}^{N_\mu} w_i^\mu \left| \hat{y}_i^{\alpha(\mu)} - \hat{y}_i^{\beta(\mu)} \right|^2 \tag{2}$$

The optimal set of transformations is the set that gives the global minimum of the objective function.

It should be noted that the set will always have a global degeneracy. The cost function is unchanged if the same rigid body transform is applied simultaneously to all M views. For this reason we may arbitrarily set the transform for view 1 to the identity transform with no loss of generality.

We do not do so during the computation because it is preferable to maintain symmetry. This means that all views may be programmed in the same way. When the computation is complete all views are transformed so that view 1 is transformed by the identity transform.

4. Finding the optimal transform

In this section we present a method for finding a local minimum of the cost function. Considerable insight is gained by creating a fictitious physical system in which the laws of classical mechanics are obeyed. In effect we will present a solution based on a gradient descent method, however useful insights may be gained by the physical analogy.

We begin by considering the problem of two views and one correspondence set, i.e. $\alpha = 1$ and $\beta = 2$. The cost we wish to minimize is

$$E[R^1, R^2, T^1, T^2] = \sum_{i}^{N} w_i \left| \hat{y}_i^\alpha - \hat{y}_i^\beta \right|^2 \tag{3}$$

For each of the M views we choose an arbitrary point in space which we will call the center of mass and denote x_{CM}. We defer discussion on how to choose x_{CM} until later. We can now define points relative to the center of mass of a view before and after applying equation (1). [The hat ^ is used to distinguish points with and without the CM subtracted.]

$$\hat{y}_i = y_i + y_{CM} \tag{4}$$

$$\hat{x}_i = x_i + x_{CM} \tag{5}$$

We now define some averaged (mean) quantities. The mean value of $\hat{x}_i$ is denoted $\langle \hat{x} \rangle$ and given by

$$\langle \hat{x} \rangle = \frac{\sum_i w_i \hat{x}_i}{\sum_i w_i} \tag{6}$$

Other averaged quantities are defined in the same way. For convenience we define $W = \sum_i w_i$. We note now a useful formula which follows from the above definition.

$$\langle x \rangle = \langle \hat{x} - x_{CM} \rangle = \langle \hat{x} \rangle - x_{CM} \tag{7}$$

In order to make it quite clear we note here that there are M centers of mass, but $2P$ mean values $\langle x \rangle$. This is because each view has a *fixed* center of mass but the $\langle x \rangle$ are computed based on the actual points and weights for a given correspondence set.

The pairwise point cost is given by the squared Euclidean distance. This is similar to the harmonic oscillator potential in physics. The potential energy stored in a spring is the square of its displacement from the resting position. The potential gives rise to a force linearly proportional to displacement. The force may be viewed as attached to a rigid object (view) at $\hat{y}$. In figure 1 we attempt to illustrate the model by showing 3 objects with fixed points joined by springs. This is the dynamical system we base the model on. In the figure there are 3 views and 3 correspondence sets.

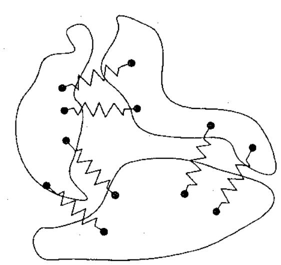

Figure 1. The physical model

In physics a dynamical system with friction will evolve towards states with lower potential energy and eventually come to rest in a local minimum of potential energy. This is very similar to the method of gradient descent in optimization where the potential energy plays the role of the cost or objective function. The physical forces play the role of the gradients which specify the update direction.

The force on view α due to the correspondence set may be computed as

$$\begin{aligned} F^\alpha = -\sum_i \frac{\partial E}{\partial \hat{y}_i^\alpha} &= -2W \left[R^\alpha \langle x^\alpha \rangle + R^\alpha x_{CM}^\alpha + T^\alpha \right. \\ & \left. -R^\beta \langle x^\beta \rangle - R^\beta x_{CM}^\beta - T^\beta \right] \end{aligned} \tag{8}$$

A similar expression is obtained for F^β. The torque on view α, τ^α, around the center of mass may be computed as

$$\begin{aligned} \tau^\alpha &= -\sum_i y_i^\alpha \times \frac{\partial E}{\partial \hat{y}_i^\alpha} \\ &= -2W[R^\alpha \langle x^\alpha \rangle] \times \left[R^\alpha x_{CM}^\alpha + T^\alpha - R^\beta x_{CM}^\beta - T^\beta \right] \\ & +2\langle R^\alpha x^\alpha \times R^\beta x^\beta \rangle \end{aligned} \tag{9}$$

Here we have used the fact that $\langle x^\alpha \times x^\alpha \rangle = 0$.

It is possible to reduce the complexity of computing the force and the torque from $O(N)$ to $O(N^0)$ by precomputing quantities such as the $\langle x^\alpha \rangle$ and storing them. The term $\langle R^\alpha x^\alpha \times R^\beta x^\beta \rangle$ is slightly more difficult to precompute. It may be expressed in terms of the 3×3 correlation matrix.

It is now straightforward to generalize the 2 view case to M views. From one correspondence set we obtain a force and torque for each of the two participating views. By summing over all P correspondence sets we can obtain a total force $F_{tot}^{\alpha'}$ and torque $\tau_{tot}^{\alpha'}$ for each of the M views.

$$F_{tot}^{\alpha'} = \sum_\mu \delta(\alpha', \alpha(\mu)) F^{\alpha(\mu)} + \sum_\mu \delta(\alpha', \beta(\mu)) F^{\beta(\mu)} \tag{10}$$

A similar expression holds for the torque. Now that we have the total force and torque we can consider the dynamical equations of motion. We seek a system that moves towards a potential minimum. The equations that we choose are the friction dominated equations of motion, namely

$$\gamma \frac{dy_{CM}^\alpha}{dt} = F_{tot}^\alpha \tag{11}$$

$$\Gamma \omega^\alpha = \tau_{tot}^\alpha \tag{12}$$

The parameters γ and Γ resemble the mass and moment of inertia. In this system they are in fact the drag and rotational drag coefficients. ω is the angular velocity, i.e. the time rate of change of orientation. [The angular velocity forms a vector space whereas the orientation does not.]

We note that Newtons laws are usually second order and these equations are first order. This is because the mass terms have been dropped leaving only the first order friction terms. Hence we describe the equations as friction dominated.

This system of equations can be solved by the Euler method. Each view must be translated by $F_{tot}^\alpha \Delta t$ and rotated by $R_{\omega\Delta t}^\alpha$ which is obtained from $\omega \Delta t$ in a standard way. The update equation for $T^\alpha(t)$ is complicated because the rotation is around the center of mass, not the origin.

$$T^\alpha(t+\Delta t) = R_{\omega\Delta t}^\alpha [T^\alpha(t) - y_{CM}^\alpha] + y_{CM}^\alpha + F_{tot}^\alpha \Delta t \tag{13}$$

$$R^\alpha(t + \Delta t) = R^\alpha_{\omega \Delta t} R^\alpha(t) \tag{14}$$

A simple quality controlled Euler method can solve this dynamical system with adaptive stepsize. It is guaranteed to converge to a local minimum.

The method must be started at some initial guess. In the results section we show convergence from rather crude initial guesses. To save time the pairwise method may be used to supply quite good initial guesses.

There are some arbitrary parameters to be chosen. These parameters do not affect the solution but can have a strong effect on the convergence to the solution. The drag may be set to $\gamma = 1$ with no loss of generality. Our experiments suggest that correct choice of the rotational drag Γ and the center of mass for each view is very important. The center of mass for a view should be chosen as the centroid of all the data originating from that view in any correspondence set. The rotational drag should be chosen as $\Gamma = 0.5\gamma R^2$ where R is a distance. The distance is selected as follows. For each view we compute the rms distance of points from the CM. R is then set to the largest rms distance of any view.

Choice of these parameters is important for the following reason. Consider the potential $E = ax^2 + by^2$. When a and b are very different in size gradient descent is poor because it tends to zigzag down the valley in an inefficient way. Rescaling so that $a \simeq b$ does not affect the position of the minimum but does affect the speed of gradient descent. In effect the above choices generate a reasonably shaped potential surface on which to perform gradient descent.

5. Results

In the first example we show the convergence properties of the algorithm by displaying a graph of (pseudo)time against rms residual r defined by

$$\left(r[R^1..R^M, T^1..T^M]\right)^2 = \frac{\sum_\mu \sum_i^{N_\mu} w_i^\mu \left| \hat{y}_i^{\alpha(\mu)} - \hat{y}_i^{\beta(\mu)} \right|^2}{\sum_\mu \sum_i^{N_\mu} w_i^\mu} \tag{15}$$

The synthetic data set is created by selecting a random set of 50 points on the surface of a bust of Beethoven. We do not add noise. [In figure 4 we show the Beethoven bust and the icosahedron that we will use later.] We then select 6 subsets corresponding to views taken from the 6 directions $+e_x$, $+e_y$, $+e_z$, $-e_x$, $-e_y$, $-e_z$. For example view 1 contains only points visible from the $+e_x$ direction. Thus $M = 6$. Each view is then rotated by 10 degrees around a random axis and translated by a random distance of about 2 units in each direction. The bust is of size about 10 units. There are 15 distinct pairs of views but only 12 correspondence sets (P=12) since views from opposite directions have no common points.

The first curve in figure 2 shows the residual error r as a function of time t for the case $M = 2$ (where we discard 4 views and have $P = 1$) and $M = 6$. The algorithm is terminated after 41 (quality controlled) iterations for $M = 2$ and 57 iterations for $M = 6$. It is clear that although the 6 view problem is considerably more difficult than the 2 view problem it needs only $\sim 50\%$ more iterations. The results are good, the cost has been reduced uniformly over 12 orders of magnitude.

In this problem ground truth is available and so we can compute the error from the true value for the translation and rotation. It is about 10^{-12} for distance and 10^{-11} degrees for rotation.

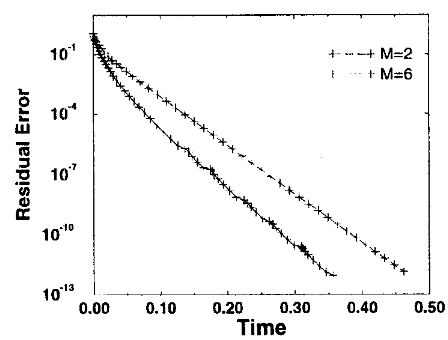

Figure 2. Convergence properties

In summary we have demonstrated fast convergence to results of high accuracy. Typical vision users will probably be satisfied with 6 decimal places or less. In addition the algorithm may be started with the results of the pairwise fittings. In practical situations no more than a few iterations may be necessary.

Whether or not multiview registration will offer significant improvement over a binary approach will depend on the details of the data in question. It is easy to construct examples where there is no advantage in using a multiview approach (e.g. no noise!). Next we present quantitative results for an example where there is a significant difference. We select 60 points at random on a icosahedron inscribed in a sphere of unit radius. We then obtain 3 views each from directions in the xy plane at 120° intervals. The directions are $(1/2, \sqrt{3}/2, 0)$, $(1/2, -\sqrt{3}/2, 0)$ and $(-1, 0, 0)$. There are 3 correspondences sets, namely views 1-2, 1-3, 2-3, and we load each set with those points common to both views. The pairwise approach can only use two correspondence sets.

We then add Gaussian noise of zero mean and variance σ to each coordinate of each point in each view, and solve for the optimal transformation. In figure 2 we show the residual plotted against σ. The residual is the full residual summed over the 3 correspondences sets in each case. The pairwise method minimizes over sets 1-2 and 1-3, whereas the iterative method minimizes over all 3 sets. The improvement is consistently larger than a factor of 2 which can be significant. Naturally the additional errors in the pairwise method will be found mainly in the set 2-3 which is not used in the minimization.

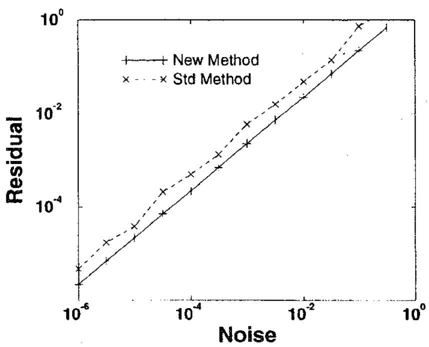

Figure 3. A graph of Residual r vs Data Noise σ comparing the pairwise method with the iterative method

6. Discussion & Conclusion

It remains an open question whether the cost function is convex. However we do know that the case of 2 point sets is convex (with respect to proper rigid body transformations). From this we can make a deduction. If we find a local minimum and fix the pose of all views except one, say α, then the cost is a *global* minimum with respect to variations of T^α and R^α.

Figure 4. The two surfaces used in the results section

We are aware of one weakness of the algorithm. In cases of near degeneracy, such as a set of points lying very close to a straight line, we find slow convergence (although the residual error r rapidly becomes small).

We have developed a simple and fast method to solve the multiview point set registration problem. After a very brief precomputation of $O(N)$ the method is of $O(N^0)$.

We have shown an example where use of this method makes a significant improvement to the estimate of pose.

References

[1] P. Besl and N. McKay. A method for registration of 3-d shapes. *IEEE Trans. Pattern Analysis and Machine Intell.*, 14(2):239–256, 1992.

[2] H. Gagnon, M. Soucy, R. Bergevin, and D. Laurendeau. Registration of multiple range views for automatic 3-d model building. In *IEEE Conf. on Vision and Pattern Recognition*, pages 581–586, New York, USA., 1994.

[3] K. Kanatani. Analysis of 3-d rotation fitting. *IEEE Trans. Pattern Analysis and Machine Intell.*, 16(5):543–549, 1994.

[4] A. Lorusso, D. W. Eggert, and R. B. Fisher. A comparison of four algorithms for estimating 3-d rigid transformations. In *British Machine Vision Conference*, pages 237–246, Birmingham, England, 1995.

[5] M. Soucy and D. Laurendau. A general surface approach to the integration of a set of range views. *IEEE Trans. Pattern Analysis and Machine Intell.*, 17(4):344–358, 1995.

[6] G. Turk and M. Levoy. Zippered polygon meshes from range images. In *SIGGRAPH*, pages 311–318, Orlando, Florida, 1994.

[7] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *Int. Journal of Computer Vision*, 13(2):119–152, 1994.