Constructing Models of Articulating Objects: Range Data Partitioning

A. P. Ashbrook
Department of Artificial Intelligence
The University of Edinburgh
Edinburgh, Scotland
anthonya@dai.ed.ac.uk

R. B. Fisher
Department of Artificial Intelligence
The University of Edinburgh
Edinburgh, Scotland
rbf@dai.ed.ac.uk

Abstract

In this paper we consider one aspect of the problem of automatically building shape models of articulating objects from example range images. Central to the model construction problem is the registration of range data, taken from different vantage points, into a common coordinate frame. This involves determining a transformation for each set of range data which aligns overlapping surface points in the common frame. Current registration algorithms have been developed specifically for rigid objects, but it is not obvious how these can be extended to articulated or more generally deformable objects. Here, we propose that range images of articulated objects are first segmented into their rigid subcomponents. Each subcomponent can then be registered in isolation using the existing algorithms designed specifically for rigid parts and the final model formed by reassembling all of the submodels. This has motivated the development of a rigid part segmentation algorithm which is described and demonstrated here. The algorithm is currently limited to non-umbilic surfaces, but in this more restricted domain is shown to work well. Current work is investigating how the approach can be extended to all surfaces types.

1. Introduction

The widespread use of 3-dimensional shape models in a broad range of applications has motivated the development of techniques for automatic model acquisition. This not only shortens the time it takes to construct models but allows models to incorporate levels of detail which would be impractical to construct manually. To generate a complete description of the surface shape of an object, multiple range measurements are usually taken from a number of vantage points. These sets of range data are then *registered* into a common coordinate frame and then used to determine the presence and position of object surfaces [3].

To register each set of range measurements into a common coordinate frame, the transformation that aligns overlapping surface points in the common frame must be determined. For a single rigid and stationary object this transformation simply describes the change in pose of one vantage point to another and can be determined by aligning overlapping surface features using their local shape. For deformable objects the problem is much more complex as each transformation not only describes the change in pose of the vantage point, but also the change in shape of the object itself. Because of this increased complexity, current systems for automatic model acquisition tend to be limited to rigid objects.

In this paper we extend the current model acquisition techniques to articulated objects and in particular the registration of range data describing articulating objects. Articulated objects are the simplest class of deformable object, and consist of a number of rigid parts which are connected by non-rigid joints [6]. The final model should not only describe the surface geometry of the object but should also contain a kinematic description of the relative movement of the different parts. Examples of articulated objects that we may eventually wish to acquire automatically include robot arms, machine assemblies and, at least to first order, human beings and animals. A number of researchers have already used articulated models in applications such as the recognition of simple articulated objects [5] and to analyse articulated motion [8]. In these examples, however, the models have been constructed manually beforehand. The problem that we are interested in is the automatic construction of these models from range data.

Unfortunately, current registration algorithms cannot be used to register articulated range data directly because of the change in the object's shape between frames. To avoid this problem we propose that each set of range data is first partitioned into its rigid subcomponents, allowing the range data of each subcomponent to be registered in isolation using current registration algorithms. The final shape model can then be produced by reassembling all of its subcomponent models. In some related work on motion analysis, Goldgof [4] developed an algorithm for partitioning articulated objects into their rigid subcomponents by identifying the object's joints. He observed that the curvature on the surface of the rigid subcomponents remains unchanged during an articulated motion but that the curvature at the actual joints does change. The object can easily be partitioned into its subcomponents by removing surface points whose curvature changes over time. The technique works well in principle and has been demonstrated on simulated data but does not provide a general solution. A particular limitation of the technique is the need for the joint between subcomponents to be fully visable in the range data. Frequently in real situations joints are either partially or even fully occulded.

The algorithm we have developed processes a pair of range images at a time and partitions each of them into N subimages, where N is the number of independently moving, rigid subcomponents that are present in the data. This processing is carried out in two distinct stages. In the first stage, the N rigid transformations that align each subcomponent in the first image with the corresponding subcomponent in the second are estimated. In the second stage, the movement of each surface data point between the two images is compared with the N estimated transformations and grouped with any they agree with. This results in the partitioning we require.

In the next section a detailed description of our rigid part segmentation algorithm is presented. This is followed by a demonstration of the algorithm applied to some real range image data. Finally we summarise the contribution made by this work and briefly discuss the aims of ongoing and future work.

2. Rigid Part Partitioning

Given a pair of range images describing an articulated object, our objective is to partition those images into N subimages that correspond to the object's N rigid parts. We begin by observing that corresponding surface points that lie on the same subcomponent will be aligned by the same rigid transformation, whilst corresponding surface points that lie on different subcomponents will be aligned by a different rigid transformation. Consider the hypothetical data in Figure 1 for example. Corresponding points on part A are aligned by the transformation $\mathbf{T}_A$, whilst corresponding points on part B are aligned by $\mathbf{T}_B$. This difference provides a mechanism for distinguishing surface data from different subcomponents and is the basis of our partitioning strategy.

To implement this strategy it is necessary to find corresponding surface points in the two range images and to determine the rigid transformation that aligns them. For non-umbilic surfaces, this can be done using the magnitude

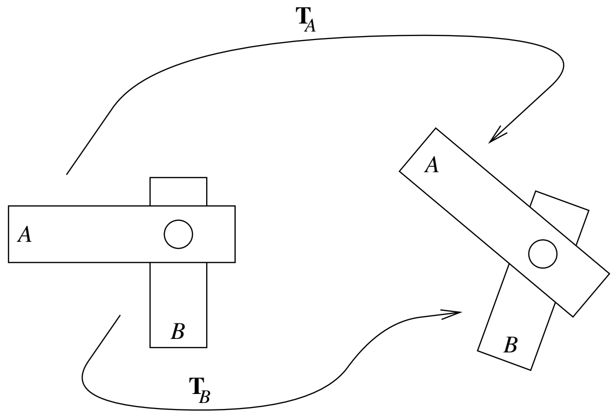


Figure 1. A hypothetical, articulated object comprising two subcomponents connected by a revolute joint. Features on the two subcomponents can be differentiated by the transformation that aligns them between frames.

and directions of the local principal curvatures at each range data point. These measurements, along with the surface normal, define the augmented Darboux frame [1] and are depicted in Figure 2. On umbilic surfaces the magnitude of the surface curvature is the same in all directions so the directions of principal curvature are not defined and the Darboux frame does not exist. To find potential correspondences between surface points in the two range images, the magnitudes of the minimum and maximum principal curvature of the surface at each point are compared. If the relative difference between both of these curvatures for a pair of surface points is within a specified level of tolerance then the surface points are considered to be a potential match. If the maximum and minimum principal curvatures at a point in the first range image are κ_{1A} and κ_{2A} respectively, and if the maximum and minimum principal curvatures at a point in the second range image are κ_{1B} and κ_{2B} respectively, then the pair of points are accepted as a potential correspondence if the following conditions are met.

$$\frac{\kappa_{1A} - \kappa_{1B}}{\frac{1}{2}(\kappa_{1A} + \kappa_{1B})} < \tau_1 \tag{1}$$

And.

$$\frac{\kappa_{2A} - \kappa_{2B}}{\frac{1}{2}(\kappa_{2A} + \kappa_{2B})} < \tau_2 \tag{2}$$

where the level of tolerance is defined by τ_1 and τ_2. The transformation, $\mathbf{t}_i$, that aligns the Darboux frames at these points is then treated as a hypothesis of the transformation that aligns one of the articulated object's subcomponents. The set of all potential surface point correspondences then provides a set of hypothetical transformations, $\{\mathbf{t}_i\}$, for aligning all of the rigid subcomponents present in the range data.

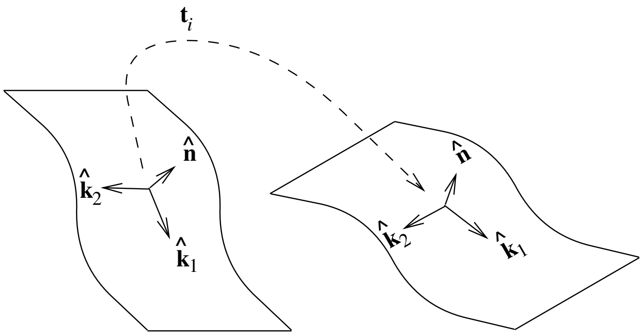


Figure 2. If the relative difference between both the maximum and minimum curvature at the two surface points is small then they are considered as a potential match. The transformation that aligns the Darboux frame defined at these points then provides a hypothesis of the transformation that aligns the surfaces.

The use of the Darboux frame for finding point correspondences and for estimating the transformations that align subcomponents does have a number of disadvantages, however. First, the Darboux frame only provides a local description of surface shape and it is likely that many different parts of a surface have the same shape locally. This results in a large number of incorrect surface point correspondences being formed. This problem is made worse by the fact that curvature measurements are sensitive to noisy data and quantisation effects, so to ensure that correct point correspondences are found it is necessary to allow a large amount of tolerance when comparing curvature. Similarly, the orientation of the Darboux frame is also sensitive to noisy data, resulting in a poor estimate of the alignment transformation for correct surface point correspondences. The requirement of non-umbilic surfaces, on which the Darboux frame is uniquely defined, is also a serious limitation, but we are currently investigating how this approach might be extended to planar and spherical surfaces.

2.1. Estimating the Alignment Transformations

We are now left with the problem of estimating the N rigid transformations, $\{\hat{\mathbf{T}}_j\}$, that align each of the subcomponents present in the pair of range images from the set of hypotheses, $\{\mathbf{t}_i\}$. This is a relatively difficult estimation problem because the number of rigid subcomponents is unknown, the correct hypotheses are subject to a relatively large error and because many of the hypotheses are misleading because of correspondence errors. Given the difficulty of this problem we have employed an approach based on the probabilistic Hough transform [9] which is known to be a robust estimator.

This use of the probabilistic Hough transform is best described by first considering a simple case in which the range images only contain a single rigid object which has undergone a transformation, $\mathbf{T}_1$. If we have some estimate of the error on each hypothesis we can estimate the probability of measuring the set of hypothetical transformations for any possible transformation $\mathbf{T}$. To make this calculation tractable we make the assumption that each hypothesis is independent. This probability can then be expressed as:

$$P(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_n|\mathbf{T}) = \prod_i P(\mathbf{t}_i|\mathbf{T}) \qquad (3)$$

where the probability $P(\mathbf{t}_i|\mathbf{T})$ describes the error on each hypothesis and $\prod_i P(\mathbf{t}_i|\mathbf{T})$ is called the likelihood function. Intuitively, a good estimate, $\hat{\mathbf{T}}_1$, of the rigid transformation $\mathbf{T}_1$, that accounts for the set of measured hypotheses, is obtained by finding the value of $\mathbf{T}$ at which this function is a maximum. This is the *most likely* value for $\mathbf{T}_1$ given the set of hypotheses.

$$\hat{\mathbf{T}}_1 = \mathrm{Max} \prod_i P(\mathbf{t}_i|\mathbf{T}) \qquad (4)$$

In the probabilistic Hough transform algorithm the likelihood function is maximised by quantising its allowable domain and exhaustively evaluating it at every interval. In practice the logarithm of the likelihood function is evaluated for efficiency as this only requires addition operations. The probabilistic Hough transform, $H(\mathbf{T})$, is therefore defined as:

$$H(\mathbf{T}) = \sum_i \ln[P(\mathbf{t}_i|\mathbf{T})] \qquad (5)$$

For estimating alignment transformations in this application, the Hough transform represents a 6-dimensional parameter space into which entries are accumulated for each hypothesis, $\mathbf{t}_i$. In practice the 6-dimensional parameter space is separated into a pair of 3-dimensional parameter spaces, representing the rotation and translation components of the estimated transformations respectively.

So far, little has been said about the hypothesis error function $P(\mathbf{t}_i|\mathbf{T})$ but the definition of this is crucial if the algorithm is to be robust to incorrect hypotheses and if it is to be able to determine multiple transformation estimates. Consider a point $\mathbf{T}$ in the parameter space that corresponds to the transformation that aligns one of the articulated objects subcomponents. If a particular hypothesis, $\mathbf{t}_i$, is a measurement of this particular transformation then $P(\mathbf{t}_i|\mathbf{T})$ simply describes the measurement error. If $\mathbf{t}_i$ is a measurement of one of the other alignment transformations or is incorrect because of a correspondence error then $\mathbf{T}$ tells us nothing about $\mathbf{t}_i$. Without further information the best we can do is say that $P(\mathbf{t}_i|\mathbf{T})$ is constant. To combine this information into a single error function, $P^*(\mathbf{t}_i|\mathbf{T})$, we describe the error within three standard deviations using a multivariate Gaussian with a uniform background probability, k, beyond this:

$$P^*(\mathbf{t}_i|\mathbf{T}) = \begin{cases} \frac{1}{2\pi|\mathbf{\Sigma_t}|^{\frac{1}{2}}} e^{-\chi_i^2/2} & \chi_i \leq 3 \\ k & \text{Otherwise} \end{cases} \tag{6}$$

where $\mathbf{\Sigma_t}$ is the covariance matrix that describes the measurement error and χ_i^2 is defined as:

$$\chi_i^2 = [\mathbf{t}_i - \mathbf{T}]^T \mathbf{\Sigma_t}^{-1} [\mathbf{t}_i - \mathbf{T}] \tag{7}$$

Having constructed the probabilistic Hough transform in this manner, the number of significant peaks will correspond to the number of subcomponents in the range data. The position of each peak in the parameter space then provides an estimate of the transformation $\hat{\mathbf{T}}_j$ that aligns the jth subcomponent in the first range image to its corresponding subcomponent in the second.

2.2. The Partitioning Strategy

The scheme outlined so far provides a robust method for determining the number of rigid subcomponents present in a pair of range images and for estimating the transformations that align each of them. This information is then used to partition each of the range images into their rigid subcomponents.

Having identified the rigid transformations that align each of the object subcomponents present in the range data, it is possible to verify each of the hypotheses, $\mathbf{t}_i$, generated earlier. For each hypothesis we calculate the probability $P(\mathbf{t}_i|\hat{\mathbf{T}}_j)$ for each of the estimated transformations $\hat{\mathbf{T}}_j$. This is the probability that the hypothesis $\mathbf{t}_i$ would have been measured if the surface points it was derived from underwent a transformation $\hat{\mathbf{T}}_j$. If this probability is large for any $\hat{\mathbf{T}}_j$ then it is probable that the surface points used to generate the hypothesis lie on the jth subcomponent. The probability of measuring a particular hypothesis $\mathbf{t}_i$ depends upon both the measurement error on each hypothesis, $P(\mathbf{t}_i|\mathbf{T})$, used earlier and the error on the transformation estimate. Because the transformation estimate is derived from numerous hypotheses, the estimate error will be significantly less than the hypothesis error and so the following approximation can be made:

$$P(\mathbf{t}_i|\hat{\mathbf{T}}_j) \approx P(\mathbf{t}_i|\mathbf{T}) \tag{8}$$

We then define the distance, d_{ij}, between the hypothesis and the estimate in standard deviations using the Mahalanobis distance metric:

$$d_{ij}^2 = [\mathbf{t}_i - \hat{\mathbf{T}}_j]^T \mathbf{\Sigma}_j^{-1} [\mathbf{t}_i - \hat{\mathbf{T}}_j] \tag{9}$$

If we then want 99.9% of the surface points to be included in our partitioning, the surface points used to generate the hypothesis $\mathbf{t}_i$ should be assigned to subcomponent j if:

$$d_{ij} \leq 3 \tag{10}$$

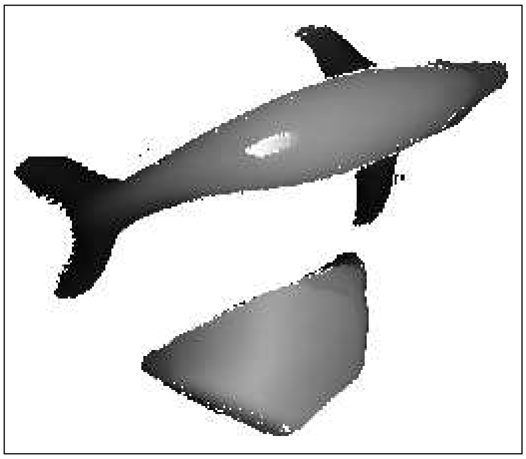
3. Experiments

Preliminary results of the rigid part partitioning strategy are presented here for two range data sets, the first presented in Figure 3 and the second in Figure 5.

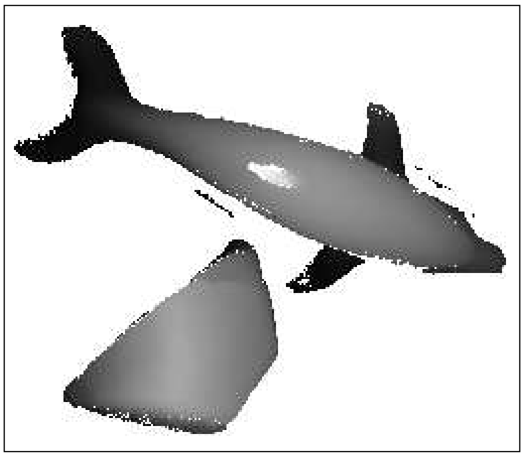
3.1. Experiment 1

Although the motivation for developing a rigid part partitioning strategy has been to isolate the rigid subcomponents of articulated objects, the technique is also suited to isolating single rigid objects which have undergone some relative motion. This is a slightly simpler problem to solve because it avoids the complication of subcomponents occluding each other, but provides a good way of testing the principle of the algorithm. In this first example, range images containing a pair of objects which have undergone a different transformation between frames have been used. See Figure 3.

Both of the Hough transform arrays used in this experiment, to represent the rotation and translation parameters, contained 100^3 bins, occupying 8Mb of memory each. The total time to run the experiment was approximately 6 hours and 20 minutes on a 50Mhz Sparc 10. Details of the five largest peaks in the probabilistic Hough transform constructed for the rotation parameters are presented in Table 1. The second column of the table presents the relative height of each peak as compared to the highest peak. The two largest peaks correspond to the two objects present in the range data and the rotation parameters associated with these peaks agree with the relative movement of the objects. It is difficult to make a quantitative analysis of this result because the objects have been placed by hand so the actual transformation of each object is not accurately known. An important observation concerns the relative magnitudes of the peaks representing the two objects and the peaks which occur by chance. The height of each peak relative to the biggest is presented in the second column of Table 1. In this example the first two peaks are well separated from the rest so that thresholding the Hough space is relatively straightforward.



(a)



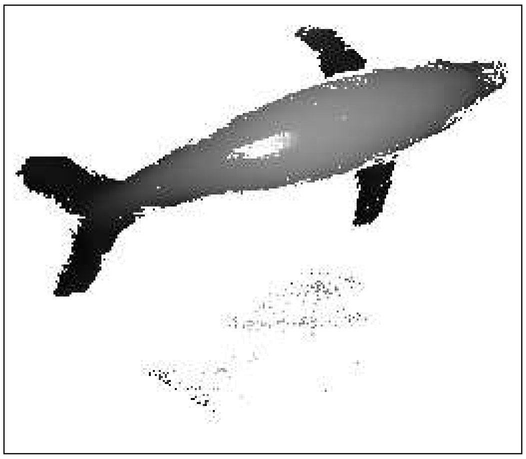
(b)

Figure 3. Each of the objects in these range images has rotated between frames. The difference in the respective transformations provides the necessary information for partitioning the range data into its rigid parts

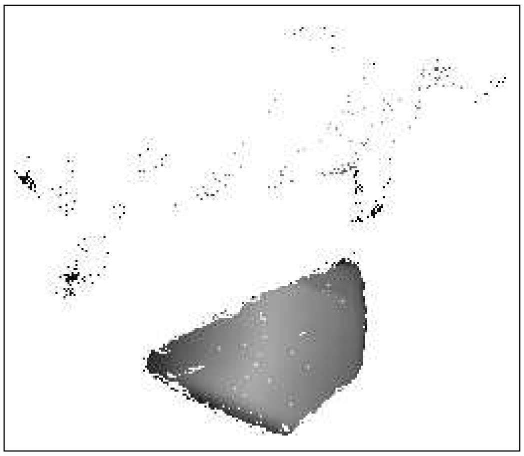
Peak Height	Rel. %	θ_z	θ_y	θ_x
98077.6	100.0	-41.4	1.8	1.8
62013.4	63.2	12.6	1.8	1.8
13600.3	13.9	84.6	-34.2	171.0
13313.4	13.5	109.7	-30.6	171.0
10944.4	11.2	-48.6	37.8	-27.0

Table 1. Details of the five largest peaks in the rotation Hough transform constructed for the range data in Figure 3.

Figure 4 presents the partitioning of the range image in Figure 3 (a) into separate rigid objects. Overall the algorithm has performed very well and the majority of surface data points have been partitioned correctly. Most of the erroneous points tend to lie close to the edge of the objects where the estimation of surface curvature is less stable. It should be possible to improve the result further by post-processing the partitioned data using morphological operators.



(a)



(b)

Figure 4. These images present the partitioning of the range image in Figure 3 (a).

3.2. Experiment 2

In this experiment a mock-up of a simple articulated joint is used to provide a more comprehensive test of the partitioning algorithm. The joint is formed by placing a conical object on a cylindrical surface and articulated motion is simulated by moving the cone over the surface. Figure 5 presents this joint with the cone at two different inclinations. In this example the partitioning is more complicated because the rotation is no longer in the image plane and parts of the subcomponents are occluded.

The same Hough transform array sizes were used in this example, but, because of the smaller number of surface points, the execution time was slightly less at 5 hours and 10 minutes. Details of the five largest peaks in the probabilistic Hough transform constructed for the rotation parameters are presented in Table 2. The two largest peaks correspond to the two subcomponents present in the range data and the rotation parameters associated with these peaks agree with their relative movement.

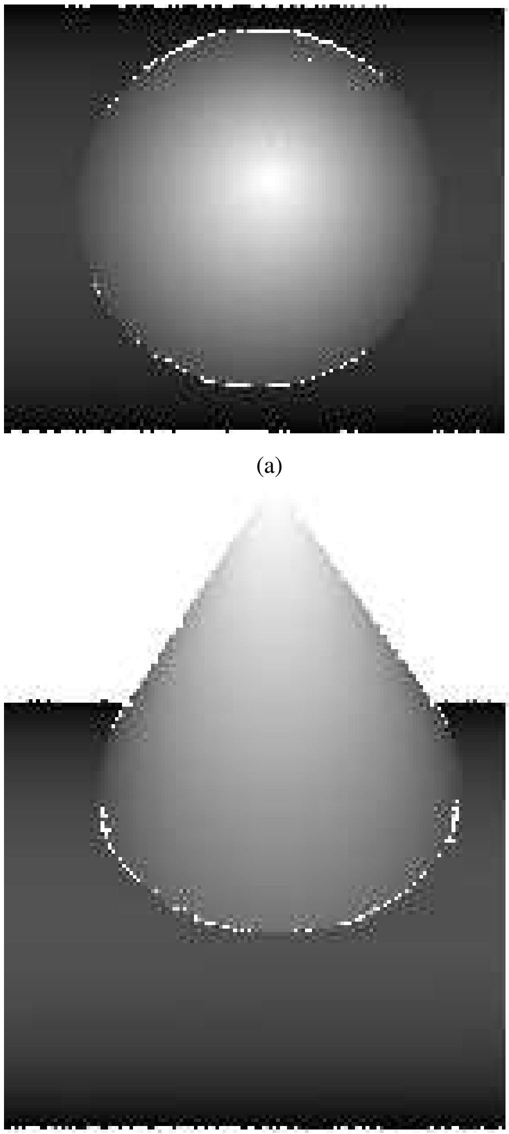
Peak Height	Rel. %	θ_z	θ_y	θ_x
56234.1	100.0	1.8	1.8	55.2
33403.1	59.4	1.8	3.4	1.8
17770.0	31.6	24.5	12.4	64.2
9897.2	17.6	-98.1	17.9	17.5
4048.9	7.2	-45.3	103.5	-34.1

Table 2. Details of the five largest peaks in the rotation Hough transform constructed for the range data in Figure 5.

Figure 6 presents the partitioning of the range image in Figure 5 (a) into its rigid subcomponents. Again the algorithm has performed well, although, in each of the partitioned subimages, a distinct region of the surface data that we would expect to be present is actually missing. The reason for this phenomenon is that the missing surface points are not present in both of the range images so it is impossible to determine which subcomponent they belong to. This effect has both an advantage and a disadvantage. The disadvantage is that to obtain a complete description of the surface of an articulated object more example views are required. The advantage is that registration of this surface data is likely to be more reliable because the registration algorithm [2] works best when the data sets to be registered contain the same surface points.

4. Conclusions and Future Work

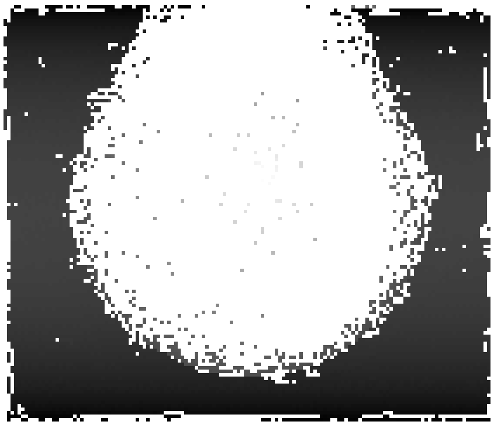
The number of applications for 3-dimensional shape models is steadily growing and the need for more complex models and a reduction in the time to construct models has motivated the development of techniques for automatic model construction. Central to the problem of automatic model construction is the registration of different views of



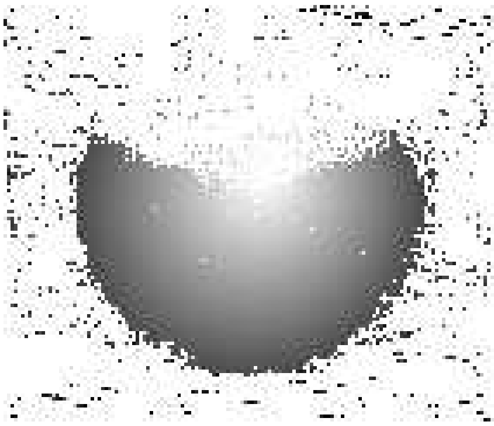
(a)

(b)

Figure 5. The conical subcomponent has rotated between frames. The difference in the respective transformations provides the necessary information for partitioning the range data into its rigid parts.



(a)



(b)

Figure 6. These images present the partitioning of the range image in Figure 5 (a).

an object into a single coordinate frame. Although this registration process is relatively straightforward for isolated, rigid objects, extending it to deformable objects has proved to be more difficult.

To avoid this registration problem we have proposed that views of articulated objects are first partitioned into their rigid subcomponents. This then allows views of each rigid subcomponent to be registered in isolation using existing algorithms and the final, articulated model constructed by assembling each of the modelled parts.

Although the idea of partitioning articulated data into rigid subsets has been used in the field of motion analysis the algorithms used are not suitable for genreral purpose model construction. The has motivated the development of a new algorithm. This algorithm has been demonstrated on real range data and appears to work adequately.

So that the algorithm may be applied more generally, we are currently investigating ways for extending it to planar and spherical surfaces. The problem with these types of surface is that the Darboux frame is not fully constrained as it can rotate about the normal vector on these surface types. In principle it is possible to use these weaker constaints. Instead of voting for a point in the parameter space we end up with a plane of votes but a number of these planes for different surface points should intersect at the correct solution.

We are also looking at the use of morphological operators for post-processing the partitioned data to obtain better results. One possible criticism of the technique is the large amount of storage and computational effort need for building and searching the probabilistic Hough transform, although this partitioning process is only performed once when creating new models. We intend to address this particular problem by employing a hierarchical strategy [7].

5. Acknowledgements

The work presented in this paper was funded by a UK EPSRC grant GR/H86905.

References

[1] do Carmo, M., “Differential Geometry of Curves and Surfaces”, Englewood Cliffs, Prentice Hall, 1976.

[2] Eggert, D., Fitzgibbon, A. W. and Fisher, R. B., “Simultaneous Registration of Multiple Range Views for use in Reverse Engineering”, Proc. ICPR, pp243-247, 1996.

[3] Fisher, R. B., Fitzgibbon, A. W. and Eggert, D., “Extracting Surface Patches from Complete Range Descriptions”, To be presented at the International Conference on Recent Advances in 3D Imaging and Modeling, Ottawa, 1997.

[4] Goldgof, D. B., Lee, H., and Huang, T. S., "Motion Analysis of Nonrigid Surfaces", Proc. CVPR88, pp375-380, 1988.

[5] Grimson, W. E. L., "Object Recognition by Computer", MIT Press, 1990.

[6] Kambhamettue, C. and Goldgof, D. B., "Nonrigid Motion Analysis", Handbook of Pattern Recognition and Image Processing: Computer Vision, Ed. Young, T. Y., Chapter 11, 1994.

[7] Li. H., Lavin, M. A. and Le Master, R. J., "Fast Hough Transform: A Hierarchical Approach", Computer Vision, Graphics & Image Processing, Vol. 36, pp139-161, 1986.

[8] O'Rourke, J. and Badler, N. I., "Model-Based Image Analysis of Human Motion using Constraint Propagation", IEEE PAMI, 2(6), pp522-536, 1980.

[9] Stephens, R. S., "A Probabilistic Approach to the Hough Transform", Proc. BMVC90, pp55-59, 1990.