Segmentation of Range Data for the Automatic Construction of Models of Articulated Objects

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 the problem of automatically constructing geometric models of articulated objects from example range data. The problem of automatic model construction has been investigated in some depth for rigid objects but these do not extend easily to the articulated case. The problem arises because of the need to register surface measurements taken from different viewpoints into a common reference frame. Registration algorithms generally assume that an object does not change shape from one view to the next but when building a model of an articulated object it is necessary for the modes of articulation to be present in the example data. To avoid this problem we propose that raw surface data of articulated objects is first segmented into rigid subsets which correspond to rigid subcomponents of the object. This allows a model of each subcomponent to be constructed using the conventional approach and a final, articulated model to be constructed by assembling each of the subcomponent models.

1 Introduction

The ability to automatically acquire geometric models from example objects is useful in a growing number of application areas. In the field of computer graphics, the need for improvements in realism requires more complex models, but manual model construction is time-consuming and difficult. In industrial settings it is useful to capture the geometry of existing parts either for the purpose of inspection or to enable exact replicas to be manufactured automatically. For reasons which will be clarified shortly, current techniques are generally limited to constructing models of single rigid objects. In this paper we suggest how these algorithms might be augmented to allow the automatic construction of articulated objects, increasing the scope of this technology. For clarity, articulated objects are those objects consisting of a number of rigid parts which are connected by non-rigid joints [1].

The established approach for automatic model construction begins by taking surface measurements from a number of viewpoints so that all of the object's surface is captured. Typically this will be done with a range finder such as a laser striper or binocular vision system. Either one of two different approaches can then be taken. In the first approach surface primitives are fitted to the raw data in each of the views of the object, and then the different views are *registered* by aligning similar primitives. The second approach registers the raw data initially using local surface shape, see for example [2] or [3], and then surface primitives are fitted directly to the registered surface data [4]. The second approach is favoured because it makes maximum use of the raw data when surface fitting and avoids the problem of having to piece together possibly fragmented surfaces from different viewpoints which are not perfectly aligned.

Whichever approach is taken, the registration process assumes that the shape of the object does not change as the surface data is acquired. If, however, we wish to automatically capture the geometry and kinematics of an articulated object then the object's shape must change from example to example. This means that current registration algorithms cannot be used directly. Rather than developing new registration algorithms we propose here that the raw measurement data is segmented into *rigid subsets* each corresponding to a rigid subcomponent of the object. This will enable models of each subcomponent to be constructed independently using existing technology and a final, articulated model to be formed by assembling each of the subcomponents.

The algorithm we have developed processes a pair of range images at a time and segments each of them into N subimages, where N is the number of independently moving, rigid subcomponents which 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 segmentation 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. The algorithms developed are also summarised in the form of pseudo-code in appendices at the end of the paper.

2 Rigid Part Segmentation

Given a pair of range images describing an articulated object, we want to segment those images into N subimages which correspond to the objects 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 different rigid transformations. 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 segmentation strategy.
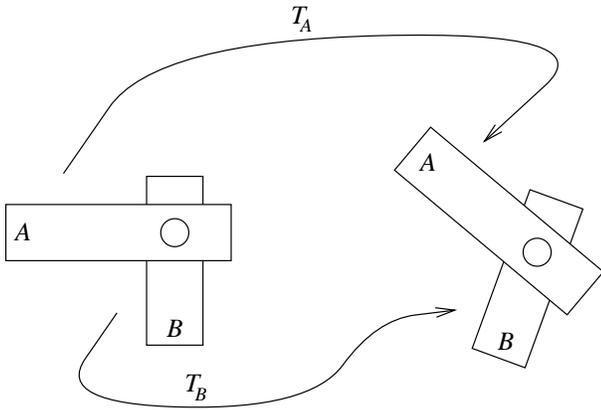

Figure 1. A hypothetical, articulated object comprising two subcomponents connected by a revolute joint. Features on the two subcomponents can be differentiated by the transformation that aligns them between frames.

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-developable surfaces, this can be done using the magnitude and directions of the local surface curvature at each range data point. These measurements define the augmented Darboux frame [5] 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 minimum and maximum curvature of the surface at each point are compared. If the relative difference between both of these curvatures 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:

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

and

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

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 objects 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.

The use of the Darboux frame for finding point correspondences and for estimating the transformations which align subcomponents has a number of disadvantages. First, the measurement of surface curvature is 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. This results in a large number of incorrect surface correspondences being formed which complicates the segmentation process later. 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-developable surfaces, on which the Darboux frame is uniquely defined, is also a serious limitation, but we are currently investigating how
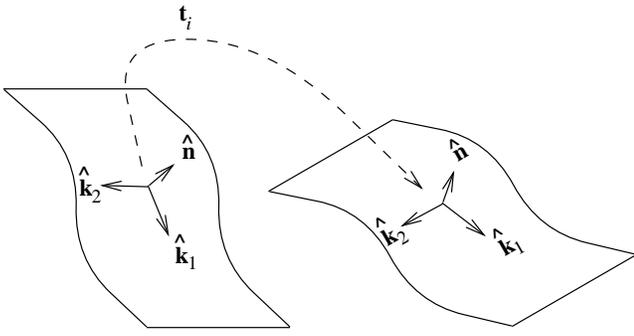
Figure 2. If the relative difference between both the maximum and minimum curvature at the two surface points is small then they are considered as a potential match. The transformation that aligns the Darboux frame defined at these points then provides a hypothesis of the transformation that aligns the surfaces.

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 [7] 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 arbitrary 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 trans-

formation $\mathbf{T}_1$, which 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.

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

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 represented a 6-dimensional parameter space into which entries are accumulated for each hypothesis, $\mathbf{t}_i$. 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 beyond this:

$$P^*(\mathbf{t}_i | \mathbf{T}) = \begin{cases} \frac{1}{8\pi^3 |\mathbf{\Sigma_t}|^{\frac{1}{2}}} e^{-\chi_i^2/2} & \chi_i \leq 3 \\ k & \text{Otherwise} \end{cases} \qquad (6)$$

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

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

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 Segmentation 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 which align each of them. This information is then used to segment 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:

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

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

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

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

$$d_{ij} \leq 3 \qquad (10)$$

3 Experiments

Preliminary results of the rigid part segmentation 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 segmentation 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.
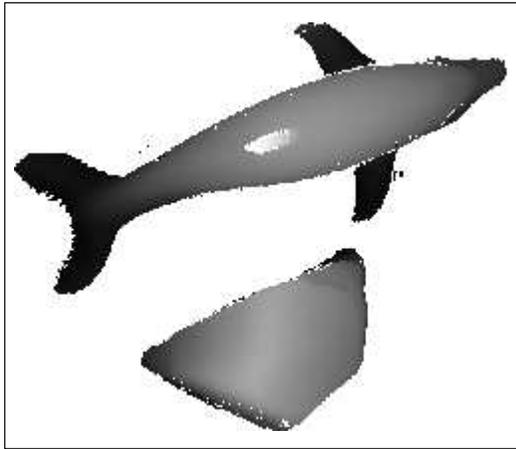
Both the rotation and translation Hough transform arrays used in this experiment contained 100^3 bins, occupying 8Mb of memory each. The total time to run the experiment was approximately 6 hours and 20 minutes on a 50Mhz Sparc 10. Details of the five largest peaks in the probabilistic Hough transform constructed for the rotation parameters are presented in Table 1. The 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.

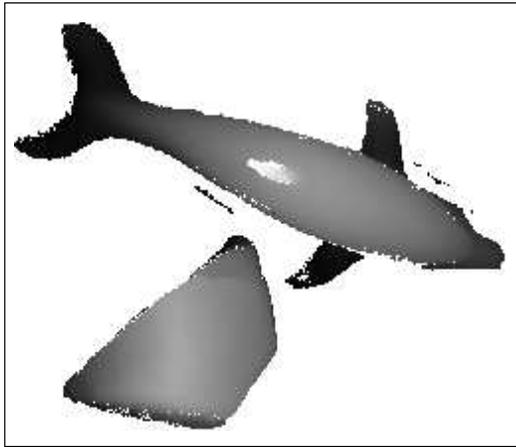| Peak Height	Peak Height %	θ_z	θ_y	θ_x
98077.6	100.0	-41.4	1.8	1.8
62013.4	63.2	12.6	1.8	1.8
13600.3	13.9	84.6	-34.2	171.0
13313.4	13.5	109.7	-30.6	171.0
10944.4	11.2	-48.6	37.8	-27.0

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 segmentation 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 segmented 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 postprocessing the segmented data using morphological operat-
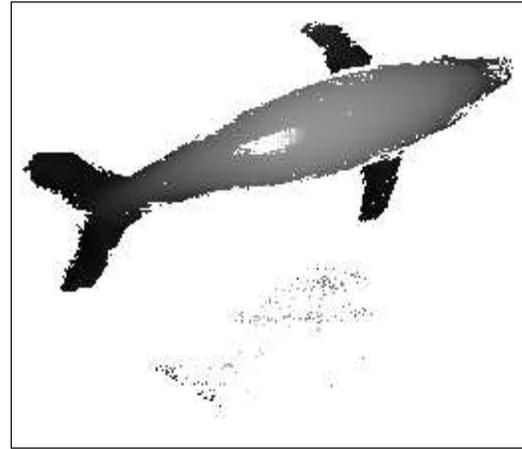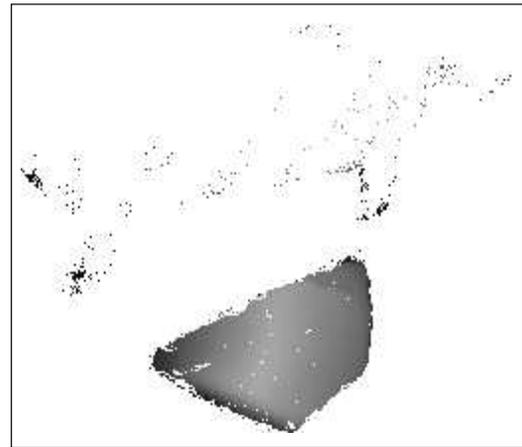
ors.

(a)

(b)

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

(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 segmenting the range data into its rigid parts

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 segmentation 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 segmentation 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
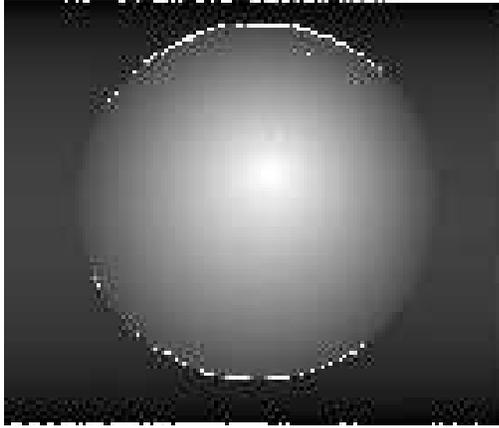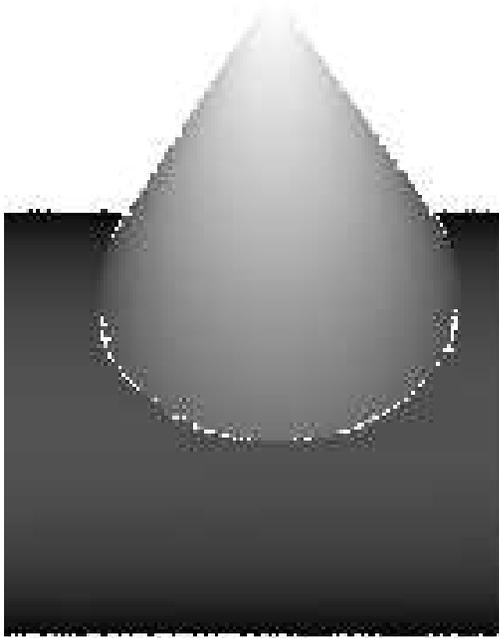
points, the execution time was slightly less at 5 hours and 10 minutes. Details of the five largest peaks in the probabilistic Hough transform constructed for the rotation parameters are presented in Table 2. The two largest peaks correspond to the two subcomponents present in the range data and the rotation parameters associated with these peaks agree with their relative movement.

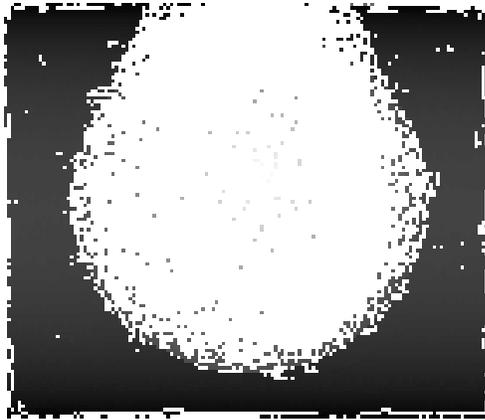| Peak Height	Peak Height %	θ_z	θ_y	θ_x
56234.1	100.0	1.8	1.8	55.2
33403.1	59.4	1.8	3.4	1.8
17770.0	31.6	24.5	12.4	64.2
9897.2	17.6	-98.1	17.9	17.5
4048.9	7.2	-45.3	103.5	-34.1

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 segmentation of the range image in Figure 5 (a) into its rigid subcomponents. Again the algorithm has performed well although in each of the segmented subimages a distinct region of the surface data which 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 registration algorithm [3] works best when the data sets to be registered contain the same surface points.
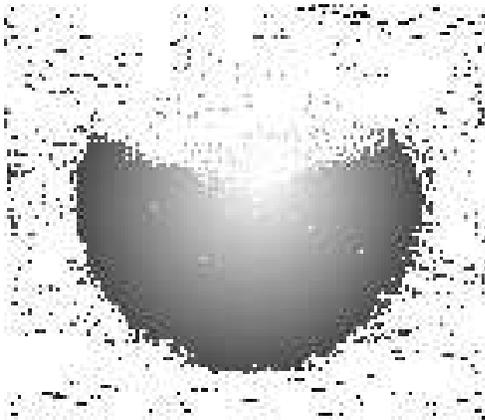
4 Organising Multiple Range Images

The algorithms presented in this paper enable the relationship between the subcomponents of an articulated object to be determined from a pair of range images. Figure 7 demonstrates how this information can be organised for multiple range images to allow reasoning about the relative motion of the different subcomponents in general. Figure 7(a) presents a connected graph which represents each of the captured range images. A node is constructed for each range image and pairs of images with overlapping data are connected by an arc. Each subcomponent is then represented by a sparse graph as in Figures 7(a) and (b). A node is constructed for each range image which contains a particular subcomponent and a pair of images containing the same subcomponent are connected by an arc. These arcs,

(a)

(b)

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

however, are directed and are labelled with the transformation which aligns the subcomponent from one image to the next. It is then straightforward to determine how each subcomponent moves relative to the others and begin to deduce how they are connected.
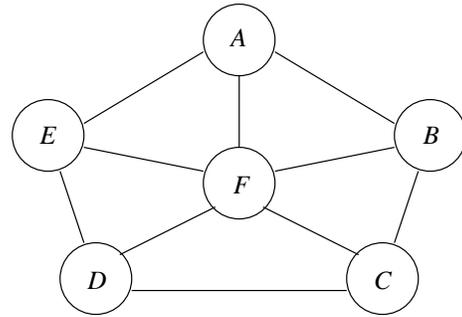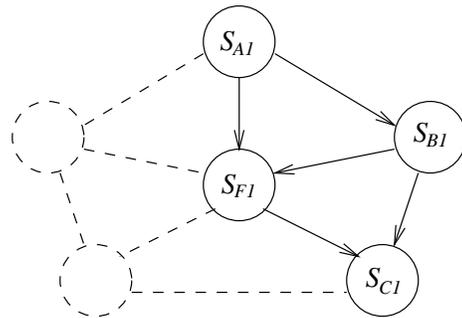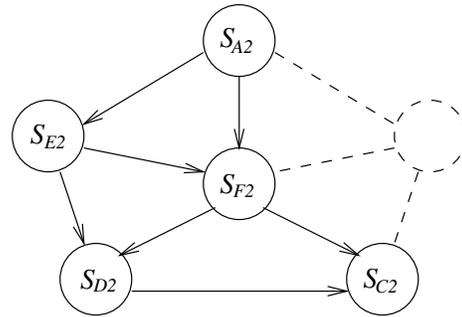

(a)

(b)

(c)

Figure 7. Organisation of multiple range images to enable reasoning about the relative motion of different subcomponents

(a)

(b)

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

5 Conclusions and Future Work

Central to the problem of automatic model construction is the registration of different views of an object into a single coordinate frame. Although this registration process is rel-

atively 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 segmented 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.

An algorithm suitable for segmenting articulated objects with non-developable surfaces into their rigid subcomponents has been developed. 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. We are also looking at the use of morphological operators for post-processing the segmented 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 segmentation process is only performed once when creating new models. We intend to address this particular problem by employing a hierarchical strategy [6].

A second problem with the Hough transform is the selection of a suitable thresholding parameter which distinguishes significant peaks from peaks which arise by chance. In the future we intend to investigate this problem further.

6. Acknowledgements

A. Estimating the Alignment Transformations

The following pseudo-code describes the algorithm used to determine the transformations that align the object subcomponents between a pair of range images.

```
for each range point a_i in image A
  construct Darboux frame F_{a_i}
  for each range point b_j in image B
    construct Darboux Frame F_{b_j}
    if curvatures at a_i and b_j are similar
      determine t_{a_i b_j} where F_{b_j} = t_{a_i b_j} F_{a_i}
      vote for t_{a_i b_j} in parameter space

for each cell in parameter space H(T)
  if H(T) > τ and is a peak
    propose T as an alignment transformation
```

B. The Segmentation Strategy

The following pseudo-code describes the algorithm used to segment each of the range images A and B into rigid subparts according to the proposed rigid transformations.

```
for each proposed transformation T_k
  for each hypothesis t_{a_i b_j}
    d = ||t_{a_i b_j} - T_i||
    if d < 3.0
      assign a_i to subimage S_{A_k}
      assign b_j to subimage S_{B_k}
```

References

[1] 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.

[2] Besl, P. J. and McKay, N. D., "A Method for Registration of 3-D Shapes", IEEE PAMI, 14(2), pp239-256, 1992.

[3] 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.

[4] Fisher, R. B., Fitzgibbon, A. W. and Eggert, D., "Extracting Surface Patches from Complete Range Descriptions", Proc. International Conference on Recent Advances in 3-D Digital Imaging and Modeling , Ottawa, 1997.

[5] M. do Carmo, "Differential Geometry of Curves and Surfaces", Englewood Cliffs, Prentice Hall, 1976.

[6] 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.

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