Automatic Acquisition of CAD Models from Multiple Range Views

A W Fitzgibbon, E Bispo,
R B Fisher, E Trucco

DAI Research Paper No. 689
March 1994

Submitted to the 10th National Conference on Manufacturing Research

AUTOMATIC ACQUISITION OF CAD MODELS FROM MULTIPLE RANGE VIEWS

A W Fitzgibbon, E Bispo, R B Fisher, E Trucco‡

Department of Artificial Intelligence
University of Edinburgh
Forrest Hill, Edinburgh EH1 2QL, Scotland
‡Department of Computing and Electrical Engineering
Heriot-Watt University
Riccarton EH14 4AS, Scotland

The project reported in this paper focuses on the automatic acquisition of CAD-like models of small industrial components from multiple range images. The goal of the target system is to generate automatically (or with minimum human intervention) object models suitable for vision tasks of industrial interest, e.g. inspection, quality control, classification, position estimation, recognition and reverse engineering.

Issues in automatic model acquisition

By an *automatic model acquisition system (AMAS)* we refer to a system capable of generating automatically a CAD-like model of an object from a number of images acquired from different viewpoints in space. There is an increasing interest for this technology, and many benefits are expected from the development of reliable AMAS, particularly for manufacturing applications: reverse engineering, product styling, NC machining, classification, recognition and inspection are only some examples.

Our work aims at developing an AMAS based on *range images*, which provide accurate, dense surface measures very efficiently. In order to acquire our own data, we have developed a low-cost, high-accuracy laser scanner capable of acquiring range images within a workspace of about $15cm^3$ with an accuracy of 0.15mm. A simple and efficient direct calibration technique has been implemented. Special consistency constraints discard most of the false measurements generated by reflective surfaces (Trucco and Fisher 1994).

The two essential issues that any model acquisition system must address are *model representation* and *view registration.* Several representations can be in principle adopted to express the models: splines, surface patches, triangulations, volumetric descriptions and finite element models are examples of possible choices. The choice of a representation is in turn linked intimately to the problem of estimating the transformation registering two successive views of the object.

In our system we use two complementary model representations, each of which implies a different solution to the registration problem: a conventional, symbolic surface patch representation (Fisher 1989, Fan 1990, Trucco and Fisher 1992), illustrated in Figure 1 (left) is combined with a B-spline model (Figure 1 (right)). The symbolic model allows fast indexing from a large database, quick pose estimation (due to the small number of corresponding features), and easy specification of inspection scripts (for example, the system can be instructed to "measure diameter of hole 2 in patch B"). On the other hand, pose estimation is limited in accuracy by the small number of correspondences and errors in the surface fitting, and provides only an incomplete surface coverage: only the most stable surface patches are retained in the model. This lack of complete data is undesirable for reverse engineering tasks, and is cured by the use of spline models. Using these models, initial pose estimates can be optimised (albeit expensively), and complete surface models easily obtained.

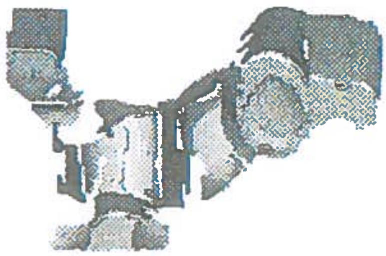
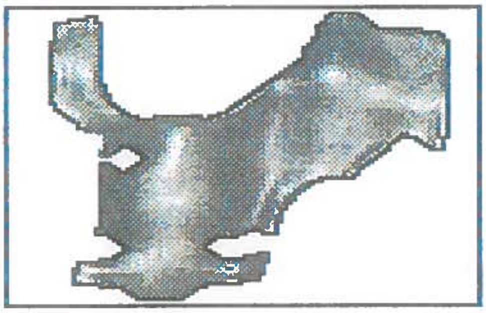

Figure 1: Example models produced by the system. On the left is a surface patch model of the truck part and its environment, on the right a coarse spline model.

In the following sections we give an overview of the computational modules of our system, namely: constructing the surface patch representation and the spline model from a single range image (view); estimating the transform between views by matching the surface patch models associated with two views; refining the estimate using the spline model; merging the views and postprocessing the current object model.

Model construction from a single range view

Building the symbolic model

Objects are initially scanned using a laser triangulation system. Depth and orientation discontinuities are detected from the raw data, and used as *a priori* weights for diffusion smoothing (Trucco and Fisher 1992). Mean and Gaussian curvatures are calculated from the smoothed image, and the data are then divided into homogeneous regions of positive, negative or zero curvature. Each region is then approximated by a viewpoint-invariant biquadratic patch (Fitzgibbon 1993), and finally expanded to include neighboring pixels which are within 3σ ($\sigma 0.15mm$) of the fitted surface. After this segmentation stage, the region boundaries are approximated by polylines and conics, and adjacent regions are intersected to produce more consistent boundaries. The resulting description is converted into a vision-oriented modeling representation, called the Suggestive Modelling System or SMS (Fisher 1987, Fitzgibbon 1992), for visualization and use in our model matching module.

Building the spline model

The spline model is constructed by laying a regular grid of control points on the image and fitting a third-order B-spline to the data. Background and noise points are removed in advance. The density of the grid is currently determined by the required accuracy—a 50 by 50 sampling allows the object in Figure 1 (right) to be approximated to within a maximum error of 0.8mm. An obvious extension is to allocate the knot points more densely at regions of high curvature, as the curvature maps are available from the segmentation process. We plan to implement this in the near future, and expect a significance increase in speed from the reduction in the number of knot points.

Estimating the transform between views

We now wish to estimate the parameters of the rigid transformation which relates two views of an object, assuming the images overlap. We start by applying the segmentation process described above to each image, thus producing two lists of surface patches. From these, an interpretation tree algorithm (Fisher 1994) finds consistent sets of corresponding pairs of surfaces. The pairs allows us to compute the 3-D pose of the object using least-squares techniques. The pose is used as an initial estimate for an iterated extended Kalman filter, which computes the uncertainty of the pose estimate.

The accuracy of view registration is within about 1σ of the noise on the range data if three or more linearly independent planar surfaces can be extracted reliably from the object (See Figure 2). Failing that, biquadratic surfaces estimated about the patch centroids are used to constrain the pose and then translation accuracy falls to about 5mm. If the pose needs to be constrained by using paired centroids, the system is open to error due to occlusion. The rotational accuracy of registration is generally within 1 degree.

Refining the inter-view transform

Given an initial pose estimate from the symbolic model matcher we can now use the spline model to refine the estimate. The pose is optimized using the Iterated Closest Point (ICP) algorithm (Besl and MacKay 1992). We have found in a 2D example that the region of convergence occupies about 25% of the space of all possible poses. In the 3D tests on the object above, a less complete investigation indicates convergence with up to 90 degrees of initial rotation error. The disadvantage of this technique is its computational complexity: for each data point, we must locate the nearest point on the model surface, then calculate the registering

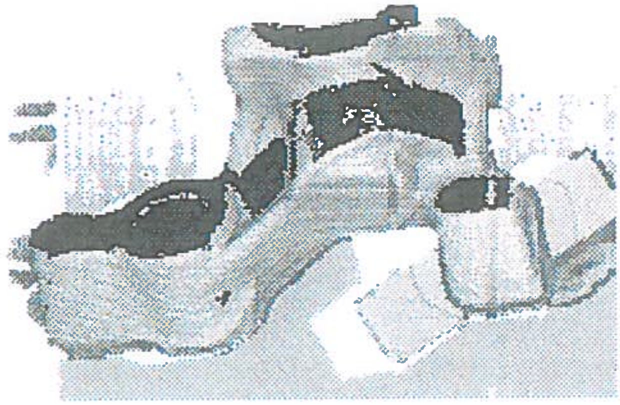

Figure 2: Surface-based matching results. Gray pixels are data, dark grey are model. The image shows a good quality match using independent planar patches and patch centroids.

transform. Locating the closest point is sped up by a combination of multigridding and subsampling the basic gradient descent algorithm. The registration accuracy is "optimal" in the sense that the noise statistics of the residuals are symmetric and white. Non-convergence does occur however, and we are currently investigating ways of further correcting for this.

View registration and model postprocessing

Final processing on the models includes merging the single-view descriptions into a single reference frame. This is done easily thanks to the SMS representation for surface patches, which separates patch descriptions into shape, extent and position. The spline models may be treated similarly, by calculating a new fitting spline for the merged sets of range data.

Discussion

We have described an AMAS capable of creating CAD-like models from range data effectively. The images need not be registered in advance, but should have some overlap between frames. The primary goal of current and future research is to investigate and improve the reliability of the symbolic model matcher, particularly under situations where occlusion and multiple objects pose difficulties. Additional topics being investigated include the extraction of features usable for manufacturing operations, e.g. holes, slots, etc. Further investigation will address the overall reliability of the system, and the improvement of the models in terms of speed and consistency.

References

Besl, P. J. and McKay, N., 1992, A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14, 239–256.
Chen, Y., and Medioni, G., 1992, Object modelling by registration of multiple range

images. Image and Vision Computing, **10**, 145–154 .
Fan, T.-J., 1990, *Describing and Recognizing 3-D Objects Using Surface Properties*, (Springer-Verlag, Berlin).
Fisher, R. B., 1987, SMS : a suggestive modelling system for object recognition. Image and Vision Computing, **5**, 98–104.
Fisher, R. B., 1989, *From Surfaces to Objects*, (John Wiley, London).
Fisher, R. B., 1994, Performance evaluation of ten variations on the interpretation-tree matching algorithm. In *Proceedings European Conf. on Computer Vision*, Stockholm, September, to appear.
Fitzgibbon, A.W, 1992, Suggestive modelling for machine vision. In *Proceedings SPIE-1830 Int. Conf. on Curves and Surfaces in Computer Vision and Graphics III*, (SPIE, Boston).
Fitzgibbon, A.W. and Fisher, R.B., 1993, Invariant fitting of arbitrary single-extremum surfaces. In J. Illingworth (ed.), *Proceedings British Machine Vision Conference*, (Springer-Verlag, Berlin) 569–578.
Trucco, E. and Fisher, R.B., 1994, Acquisition of consistent range data using local calibration. In *Proceeding IEEE Int. Conference on Robotics and Automation*, (IEEE Comp.Soc. Press, Los Alamitos), to appear.
Trucco, E. and Fisher, R.B., 1992, Computing surface-based representations from range images. In *Proceedings IEEE International Symposium on Intelligent Control*, (IEEE Control Systems Soc.), 275–280.