RECOGNITION OF COMPLEX 3-D OBJECTS FROM RANGE DATA

R. B. FISHER, A. FITZGIBBON, M. WAITE, E. TRUCCO
Machine Vision Unit
Department of Artificial Intelligence, University of Edinburgh
5 Forrest Hill, Edinburgh, EH1 2QL, Scotland

and

M. J. L. ORR
Centre for Cognitive Science, University of Edinburgh
2 Buccleuch Place, Edinburgh, EH8 9LW, Scotland

ABSTRACT

This paper describes IMAGINE, a project investigating feature-based recognition of complex 3-D objects from range data. The objects considered are bounded by surfaces of variable complexity, from planes to sculptured patches, which occur commonly in manufactured mechanical components. We introduce our current prototype, IMAGINE2, a *complete* range-based 3-D recognition system and illustrate briefly the solutions adopted in its modules, namely data acquisition, segmentation, solid object modelling, and model matching. Finally, we demonstrate the system's performance in recognizing a typical industrial component, using an automatically acquired 3-D model.

1 Introduction: the IMAGINE Project

In this paper we describe IMAGINE, a project investigating feature-based recognition of complex 3-D objects from range data. The IMAGINE project has been the *fil rouge* of the vision research of the Machine Vision Unit for many years. Although recognition has been the primary focus, many related aspects have been investigated within the IMAGINE framework. Since the last prototype, IMAGINE1, was comprehensively reported in 1989 [3], most modules have been extended and new modules added to assemble the current prototype, IMAGINE2. Today IMAGINE is a poweful environment for range-based vision and includes modules for data acquisition, segmentation, solid modeling, efficient model selection, geometric reasoning, model matching and automatic model acquisition. In this paper we shall concentrate on recognition. The modules can be used in different combinations for different tasks. In this document we describe the modules used for recognition; however, in Section 7 we shall use a 3-D object model created automatically by IMAGINE in its model acquisition configuration [1].

By object recognition we mean both *identification* (establishing the nature of an object, selecting a model) and *location* (inferring the position of the model in 3-D space from the data). IMAGINE is aimed at objects bounded by surfaces of variable complexity, from planes to sculptured patches. Surface shape is described both *qualitatively*, e.g. associating a patch to a shape class, and *quantitatively*, e.g. specifying local curvatures and normals. A small number of candidate models for recognition are efficiently invoked from a potentially very large database by a massively parallel network and passed on to feature-based matching and verification modules. An object modeling system designed specifically for vision tasks (the *Suggestive Modeling System*, SMS) allows modeling a large class of objects at different levels of details, using different classes of features (contours, surfaces, volumes) and incorporating information about visibility and sensors.

This paper offers several contributions. IMAGINE2 is an example of a *complete* range-based 3-D recognition system, ranging from data acquisition to object modelling and matching. *Full* object understanding is targeted: all visible features are matched and unsuccessful matches are explained. Statistical position estimation allows the constraining and explicit modeling of uncertainty. Finally, IMAGINE2 can represent and match generic curved surfaces (at an approximation by quadrics).

This paper describes the general architecture of IMAGINE2 (Section 2) and introduces its functionalities for recognition, namely *range data acquisition* (Section 3), *segmentation* (Section 4), *object modeling* (Section 5) and *model matching* (Section 6). We demonstrate the system's performance in recognizing two typical industrial components in Section 7, using an automatically acquired 3-D model. A few concluding remarks close the paper (Section 8).

2 The Architecture of IMAGINE2

Over the years, several IMAGINE stand-alone modules have been developed which can be configured to fit a given vision task best. In this section we describe the architecture of the recognition-oriented IMAGINE2 prototype (Figure 1). This architecture can be divided in two parts: *model-independent processing*, that occurs before the invocation process, and *model-based reasoning*, that occurs after the invocation process.

The process starts with a dense array of data acquired by a range striper [10,12]. Surface patches are segmented and extracted from the range image [14,15,16]. Surfaces are the main features adopted for recognition in IMAGINE. These patches are defined by having constant curvature sign. They are initially found based on the signs of principal curvatures [14] and then improved by robust fitting to quadratic surfaces [9]. The surface patches may be fragmented because of segmentation failures and by occlusion, so the surface patch merging process groups patches that can be well explained by a single patch.

Model matching algorithms usually have a high algorithmic complexity. To reduce the effect of this problem, we group the image surfaces into surface clusters that are likely to belong to single objects [3]. This perceptual organisation process reduces the matching complexity by eliminating unrelated features from sets of features being matched. It also allows a more focused selection of candidate models from the model base.

Model matching need not consider matching each image feature to each model feature of each model. The model invocation process selects likely models and likely features to explain each image feature and each group of image features [4]. Importantly, this process also makes model matching more efficient.

Model matching uses the candidate matches proposed by the invocation to form locally consistent groups of matches, using an *interpretation tree* [11] algorithm. The combinatorics of matching are further reduced by the use of a characteristic view representation in the object models, which cheaply encodes constraints on covisibility of features, and

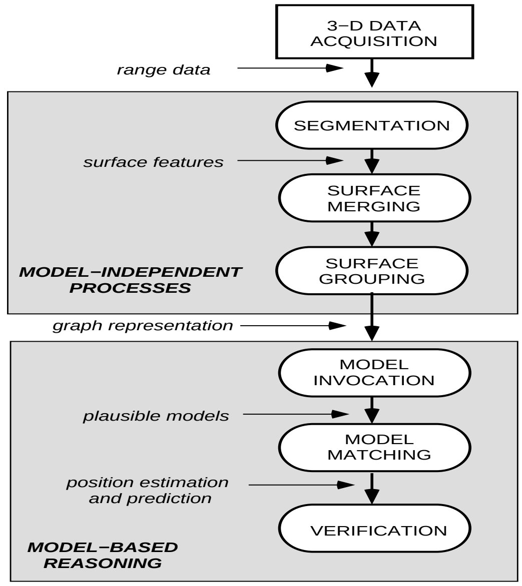

Figure 1: Architecture of IMAGINE2.

such that once a viewpoint is estimated, the matcher can determine which model features are expected to be visible. The algorithm then searches for missing features to match once a partial position has been estimated. Position estimation is based on a Kalman filter process [17], whereby geometric information is represented by a statistical distribution and individual feature matches constrain the statistical position of the whole object [13]. The advantages of such an approach are: 1) partial position estimates can be maintained, 2) likely positions of missing features and the whole object can be predicted, and 3) different types of statistical constraints can be integrated.

A final verification process ensures that model matching has not resulted in hypotheses based on coincidental surface alignments. The key constraints are surface connectivity, correct partial occlusion relationships and the location of matched features near their predicted position.

3 Range Data Acquisition

We have developed a triangulation-based range scanner [10,12] capable of acquiring images in a workspace of about $15cm^3$. The object sits on a platform moved by micro-stepper motors on a linear rail. For each position of the platform, a planar laser beam intersects the object. The resulting 3-D curve is sensed by two off-the-shelf cameras (Panasonic BL200/B) placed about one meter from the object. The acquired images are stored in a Datacube. Images are 512×512, although the sensor's workspace projects to a subwindow of the full frame. Depth measurements are first obtained independently from each camera image and subsequently merged. The use of two cameras reduces the extent of completely occluded areas and improves the depth measurements through data redundancy. Homologous points violating at least one of several consistency criteria (e.g. compatible height) are discarded [10]. A direct "black-box" technique allows the scanner to be calibrated easily, accurately and with virtually no operator intervention. Our tests indicate an accuracy at 0.15mm (repeatability about $90\mu m$) for height measurements.

4 Segmentation

Segmentation extracts a collection of homogeneous patches from the range data. A patch is a maximally connected set of surface points in which the signs of the Gaussian and mean curvature H and K do not change. Local surface shapes are divided into four main classes, planar, cylindrical, elliptic and hyperbolic according to the sign of H and K. Subclasses for concavity and convexity are also represented [14].

First the data are smoothed by solving a diffusion equation with an efficient numerical scheme [2,14]. In order to avoid the creation of noisy patches due to smoothing across discontinuities, we precompute depth and orientation discontinuities maps and use them to restrict the diffusion process to non-discontinuity points, thus avoiding the creation of spurious curved regions around discontinuity contours. An *adaptive-leakage boundary condition* is enforced at depth and orientation discontinuities [15] to limit shape distortion near boundaries. Then the signs of H and K are estimated at each nonsingular surface point. The system actually computes an exhaustive local representation (called *augmented Darboux frame*) which includes the principal directions, the normal to the surface and the principal curvatures at each surface point, but curvature signs are used as a base for segmentation because qualitative shape estimates are more reliable than quantitative surface structure estimates, and sufficient for many vision tasks. Finally, biquadratic patches are fitted to the segmented data using a viewpoint-invariant, robust fittign technique [9].

The output of the segmentation is a collection of surface patches. Each patch is associated to a qualitative shape class and to estimates of its differential structure, obtained from the biquadratic model. A *REV graph* (Region/Edge/Vertex), i.e. a winged edge graph model of the scene, expresses the structure of the scene as observed by the sensor, embedding both *intrinsic* patch properties like shape class, and *relational* patch properties like adjacency.

5 Object Modeling for 3-D Vision

IMAGINE uses the Suggestive Modelling System [7] to model 3-D objects. Objects are *assemblies* of primitive *surfaces*, forming a body centered B-rep model, a commonly available output from most CAD systems. To this model are attached *invocation properties*, and a viewer-centered representation, *viewgroups*.

The B-rep model is expressed so as to separate subcomponent *shape*, *extent* and *position* — surfaces are defined in canonical positions and paired with a reference frame transformation only when placed in an *assembly*. This encodes the inherent ambiguity arising when common subparts (a 10mm disc, say) are included in many different objects, or at several positions on the same object. Surface extent is defined by placing boundary curves onto the infinite surface, so that model verification and drawing can use boundary information while invocation and geometric reasoning need only use surface shape. Hierarchical assemblies of assemblies reduce object, viewsphere and matching complexity [5,6], and simplify the representation of articulated parts.

Invocation properties are key-value tags which store values which are expensive to derive from the model, such as surface area (patch boundaries may be arbitrarily complex), adjacency, and relative orientations. Invocation properties also encode a generic-specific type hierarchy, and an elaboration-simplification scale hierarchy.

Many vision tasks benefit from the use of a viewer-centered model database. CAD systems, however, rarely output models in this form. In addition, CAD models include surface primitives which, while part of the object's surface, will rarely be seen by the specific sensor and image processing algorithms used in the system. SMS addresses this problem by attaching to the object centered geometric model a list of *viewgroups*. This defines, for each major region on the object's visibility sphere: (1) features which must be visible if the object is unoccluded; (2) features which must not be visible from this view (for example backfacing planes); (3) hints for fast symbolic verification such as occlusion orderings, expected self occlusions, expected extremal boundaries. These groups are derived from the CAD model using a technique which decides feature saliency using an empirically derived sensor model [8].

6 Model Matching

Model-matching proceeds along two lines. First, *invocation* produces a coarse match which is fast and inexpensive [4]. This pairs every data surface with every model surface, and every data cluster with every viewgroup. Data-cluster to viewgroup pairing are assigned plausibilities based on the degree of similarity between model and data in terms of surface curvature and size. Plausibilities are assigned to data cluster to viewgroup pairings based on the plausibilities of their constituent surfaces. An iterative procedure then re-adjusts the plausibility ratings using some simple rules: for example, a surface pairing is only allowed to contribute to the most popular viewgroup in which that data surface is implicated. Eventually, a consistent state is attained and the invocation process produces a list of model surfaces compatible with each data surface, and a list of model views compatible with each data cluster.

After invocation, the second model matching process performs thorough examinations of elected pairings. For each data context, each compatible viewgroup is explored using an interpretation-tree based search technique. An attempt is made to pair every surface in the data cluster with every compatible model surface in the viewgroup. The tree search attempts to build consistent sets of these compatible surface pairings, using relative distance and orientation among the measures of consistency [6]. It is often possible for many consistent sets to be constructed. A filtering process is used to discard those sets which would be incapable of providing enough data to allow a position estimate to be constructed [17].

7 Experimental Results

This section demonstrates briefly the performance of IMAGINE2 in a recognition task using two typical industrial part. In the example involving the widget shown in

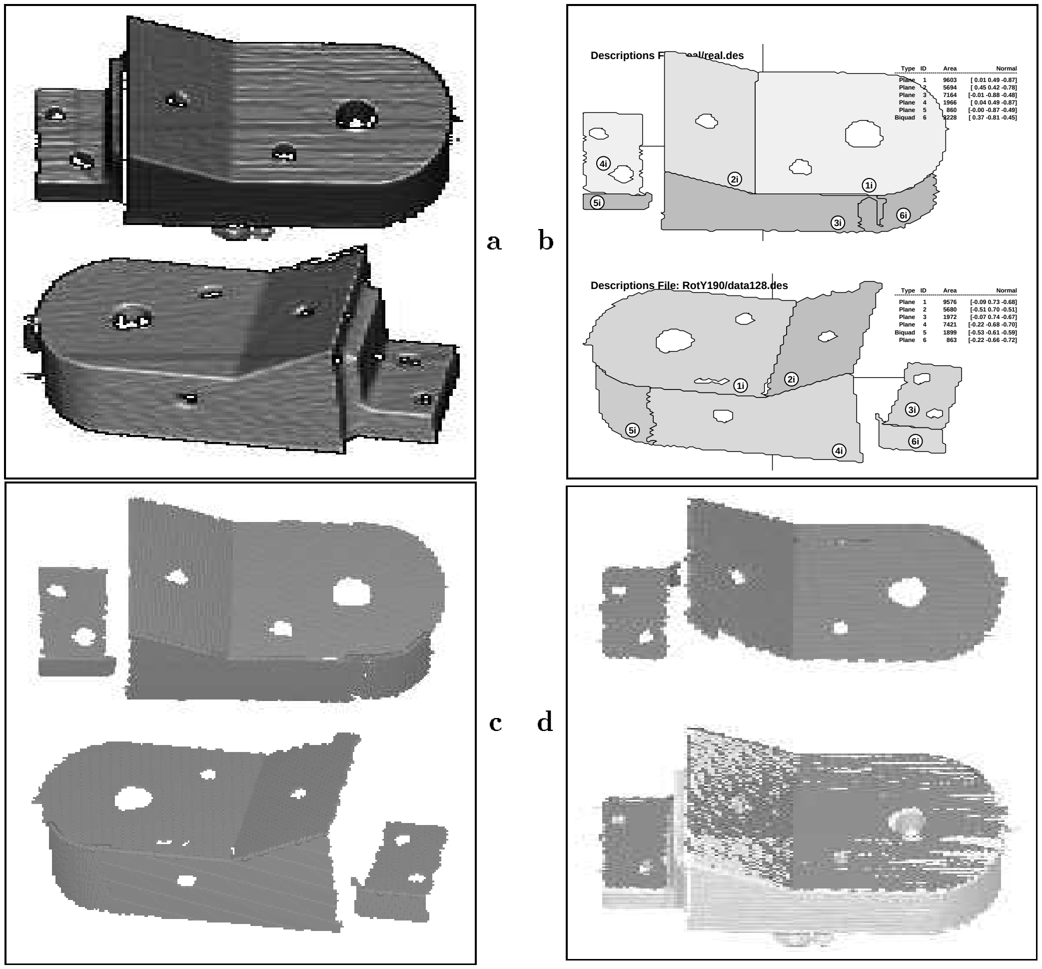

Figure 2: Stages of the recognition of a mechanical widget.

Figure 7, recognition is performed using an automatically acquired model [1] obtained from the two range views (rendered as cosine-shaded intensity images) shown in Figure 7 box (a). Box (b) shows the surface patches (cylindrical and planar) extracted by the segmentation module, and box (c) the partial SMS models built from the two views. These partial models are integrated into a single SMS model which is the one used by invocation and matching. Box (d) (top) shows the model (dark grey) superimposed to the data (light grey). The interleaving of light and dark grey demonstrates that the model surfaces are positioned within the noise error range of the data surfaces ($\sigma_{noise} = 0.15mm$). Notice that only a few data surfaces are sufficient to constrain completely the widget's position in space; more surfaces can be used to improve the accuracy of the match. Box (d) (bottom) shows the SMS model of the surfaces actually used in the match.

The second example uses a much more complex manufact, the well-knwon casting shown in Figure 8 part (a). Part (b) show the model built automatically from a different view. Part (c) demonstrates the success of the matching between range data and model. The complexity of this object depends on the number and variety of its surfaces.

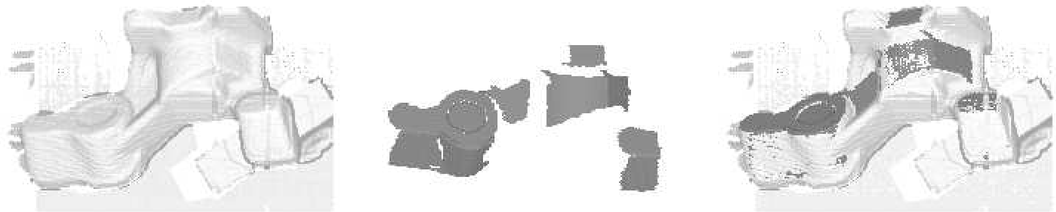

Figure 3: Recognition of a complex industrial part.

8 Discussion

This paper has briefly introduced the IMAGINE project, and illustrated a model-based 3-D object recognition system, IMAGINE2. IMAGINE2 is an example of a *complete* range-based 3-D recognition system, ranging from data acquisition to object modelling and matching. *Full* object understanding is targeted: all visible features are matched and unsuccessful matches are explained. Statistical position estimation allows the constraining and explicit modeling of uncertainty. Finally, IMAGINE2 can represent and match generic curved surface (at an approximation by biquadratics).

Ongoing research is addressing the acquisition of shape models from multiple range views, the use of biquadratic surfaces and performance assessment of IMAGINE2 with complex 3-D objects.

Acknowledgements

Parts of this work were supported by the British Science and Engineering Research Council (IED project 1551, "LAIRD") and by the Department of Trade and Industry (ACME grant GR/H/86905, "Model Acquisition from Multiple Range Views").

References

1. E.M. Bispo, A.W. Fitzgibbon, R. Fisher: *Visually Salient 3-D Model Acquisition from Range Data*, Proc. British Machine Vision Conference (BMVC-93), Surrey, September 1993.

2. L. D. Cai: *A "Small Leakage" Model for Diffusion Smoothing of Range Data.* Proc. 11th IJCAI, Detroit, 1989, pp.1585–1590.

3. R. B. Fisher: From Surfaces to Objects: Computer Vision and Three Dimensional Scene Analysis, John Wiley and Sons, Chichester, 1989.

4. R. B. Fisher: *Model Invocation for Three Dimensional Scene Understanding*, Proc. 10th Int. Joint Conf. on Artificial Intelligence, pp. 805 – 807, 1987.

5. R. B. Fisher: *Reducing viewsphere complexity*, Proc. European Conference on Artificial Intelligence (ECAI90), 1990.

6. R. B. Fisher: *Hierarchical Matching Beats the Non-Wildcard and Interpretation Tree Model Matching Algorithms*, Proc. British Machine Vision Association Conference (BMVC-92), Surrey, 1992.

7. A. W. Fitzgibbon: *Suggestive Modelling for Machine Vision*, Proc. SPIE-1830 Int. Conf. on Curves and Surfaces in Computer Vision and Graphics III, OE/Technology '92, Boston, November, 1992.

8. A. W. Fitzgibbon and R. B. Fisher: *Practical Aspect-Graph Derivation Incorporating Feature Segmentation Performance*, Proc. British Machine Vision Association Conference (BMVC-92), Leeds, 1992.

9. A.W. Fitzgibbon and R.B. Fisher: *Invariant Fitting of Arbitrary Single-Extremum Surfaces*, Proc. British Machine Vision Conference (BMVC-93), Surrey, September, 1993.

10. R.B. Fisher, D.K. Naidu and D. Singhal: *Rejection of Spurious Reflections in Structured Illumination Range Finders*, Proc. 2nd Conf. on Optical 3-D Measurement Techniques, Zurich, October, 1993.

11. Grimson, W. E. L., Lozano-Perez, T., *Model-Based Recognition and Localization from Sparse Range or Tactile Data*, International Journal of Robotics Research, Vol. 3, pp 3-35, 1984.

12. D.K. Naidu and R.B. Fisher: *A Comparative Analysis of Algorithms for Determining The Peak Posiiton of a Stripe to Sub-Pixel Accuracy*, British Machine Vision Conference BMVC91, Glasgow, September, 1991.

13. M.J.L. Orr, R.B. Fisher and J. Hallam: *Computing with Uncertainty: Intervals versus Probabilities*, British Machine Vision Conference (BMVC91), Glasgow, 1991, 351-354.

14. E. Trucco and R. B. Fisher: *Computing Surface-Based Representations from Range Images*, Proc. IEEE International Symposium on Intelligent Control (ISIC-92), Glasgow, 1992.

15. E. Trucco: *On Shape-Preserving Boundary Conditions for Diffusion Smoothing*, Proc. IEEE International Conference on Robotics and Automation, Nice, 1992.

16. E. Trucco and R.B. Fisher: *Experiments in Curvature-Based Segmentation of Range Data*, to appear in IEEE Trans. on Patt. Anal. Mach. Intell., 1993.

17. M. Waite, M.J.L. Orr, R.B. Fisher: *Statistical Partial Constraints for 3-D Model Matching and Pose Estimation Problems*, Proc. BMVC93 British Machine Vision Association Conference, Surrey, September, 1993.