3D Recognition by Parts: A Complete Solution using Parameterized Volumetric Models

DÍBIO LEANDRO BORGES[1]

Machine Vision Unit
Department of Artificial Intelligence
University of Edinburgh
5 Forrest Hill, Edinburgh EH1 2QL, United Kingdom
dibio@aifh.ed.ac.uk

Abstract. This paper presents a complete solution to the problem of recognizing 3D objects, using shape information extracted from range images, and parameterized volumetric models. The domain of the geometric shapes explored is that of complex curved objects with articulated parts, and a great deal of similarity between some of the parts. In model-based object recognition three main issues constrain the design of a complete solution: representation, feature extraction, and interpretation. We have developed an integrated approach that addresses these three issues in the context of the above mentioned domain of objects. The solution is presented here describing the main modules and how they relate to each other. Results are shown for a variety of real image data acquired in our lab.

1 Introduction

Recognition of 3D objects is an important problem that an intelligent agent has to solve for reasoning about the outside world. Developing and understanding of object recognition is one of the principal goals in the field of Computer Vision.

This work is concerned with 3D object recognition, *i.e.* identification and location, of complex curved objects with articulations. A class of natural forms is chosen to constrain the problem, thus making it possible to separate and address the issues involved. These objects are exemplified by animal shapes, however the general characteristics and complexity of these shapes are present in a wide range of other natural and man-made 3D objects.

In the model-based vision paradigm, recognition of 3D objects comprises the identification of an object present in the scene and determining its location relative to sensor position information. In order to identify what the object is, the system must have *a priori* knowledge about objects, and this knowledge is provided by building a structured geometric model in a local reference frame emphasizing particular characteristics of each object. The companion problem is to find where the object is in the scene, *i.e.* a transformation in 3D space mapping the local reference frame of the model to the data.

Three major elements can be distinguished in this paradigm: representation, feature extraction and interpretation. This work deals with all these components using range images of the 3D objects as the raw input. Figure 1 shows a functional diagram of the recognition approach presented here.

[1] Now as a Visiting Professor at CEFET-PR, Av. 7 de Setembro 3165, 80230-901 Curitiba - PR, Brazil e-mail: dibio@dainf.cefetpr.br

2 Interpretation Issues

There has not been much work in recognition of 3D articulated objects using part primitives. Previous works have addressed mainly description and symbolic identification in simple cases. [Nevatia–Binford (1977)] described a method for recognition of 3D curved objects. Description of the objects is based on finding elongated axes of the object parts using the 2D outmost boundaries of the object, and projections in eight directions of the object boundaries. [Marr–Nishihara (1978)] proposed a general method for shape recognition, in which objects are represented by a set of generalized cylinders. This set of generalized cylinders are organized as a hierarchy, where smaller parts are at the lowest levels. A process to describe animal shapes in terms of generalized cylinders was presented, in which elongated axes of the objects salient parts were found from the 2D projections of the outline of the shape. [Jain–Hoffman (1988)] described an evidence-based recognition technique for 3D objects. Surface patches are extracted from range images of objects, and a set of unary and binary properties are identified in the data. Evidence conditions with corresponding evidence weights for various objects are organized in a rulebase, and interpretation of the data is achieved by searching the rulebase to compare properties. [Fisher (1989)] presented a model invocation method which integrates evidence from object descriptions acquired from the image, and relationship evidence from class and structural associations from the model library. [Flynn–Jain (1991)] presented a constrained based search recognition system for 3D objects. Objects are described using CAD model definitions, for which special constraints are designed for the Interpretation Tree matching algorithm. Surface patches are ordered before creating the Interpretation Tree

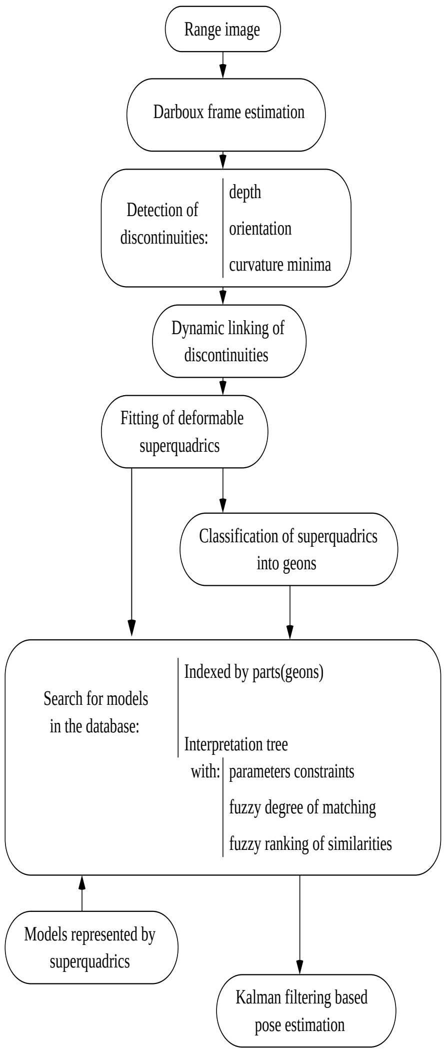


Figure 1: Functional diagram of the recognition approach developed.

according to patch area. [Grimson (1988)] proposed an extension of the Interpretation Tree matching algorithm for dealing with 2D objects, parameterized according to rotational, translation, and scaling degrees of freedom. Objects are represented by their boundaries, and matching proceeds by searching the Interpretation Tree for isolated parts.

This work deals with the 3D recognition tion issues as follows. For representation a composite description is proposed using globally deformable superquadrics and a set of volumetric primitives called geons: this description is shown to have representational and discriminative properties suitable for recognition. Feature extraction comprises a segmentation process which develops a method to extract a parts-based description of the objects as assemblies of deformable superquadrics. Discontinuity points detected from the images are linked using an "active contour" minimization technique, and deformable superquadric models are fitted to the resulting regions afterwards. Interpretation is split into three components: classification of parts, matching, and pose estimation. A Radial Basis Function [RBF] classifier algorithm is presented in order to classify the superquadrics shapes derived from the segmentation into one of twelve geon classes. The matching component is decomposed into two stages: first, an indexing scheme which makes effective use of the output of the [RBF] classifier in order to direct the search to the models which contain the parts identified. This makes the search more efficient, and with a model library that is organized in a meaningful and robust way, permits growth without compromising performance. Second, a method is proposed where the hypotheses picked from the index are searched using an Interpretation Tree algorithm combined with a quality measure to evaluate the bindings and the final valid hypotheses based on Possibility Theory, or Theory of Fuzzy Sets. The valid hypotheses ranked by the matching process are then passed to the pose estimation module. This module uses a Kalman Filter technique that includes the constraints on the articulations as perfect measurements, and as such provides a robust and generic way to estimate pose in object domains such as the one approached here.

These techniques are then combined to produce an integrated approach to the object recognition task. The next sections present an overview of the approach, results drew from the experiments performed and the main contributions of the work.

3 Representation and Feature Extraction

One of the key questions in Computer Vision research is how to segment a scene into meaningful chunks which have stable visual interpretation and can be extracted in a reliable way. This segmentation process is especially important as a first stage of a system which aims to reason about and recognize 3D objects based on the notion of prototypical parts.

To achieve the desired segmentation we developed a multi-stage approach to identifying separate parts of natural articulated objects by relying on stable estimates of the differential structure of the object's surfaces, followed by a dynamic grouping of the detected discontinuities on the surface and a fitting of deformable superquadric models to the segmented regions.

The final aim of the segmentation is to get a set of 3D deformable superquadric models, where each model of the set is related to only one part of the object. Our assumption is that three types of discontinuities (depth, concave orientation, and loci of negative minima of the Principal Directions) detected upon the local differential

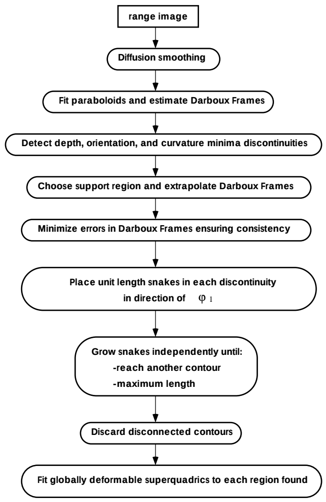


Figure 2: Functional diagram of the segmentation.

structure of the object's surface can be grouped together so as to become the boundaries of salient parts of an articulated 3D object. This grouping is formulated as an energy-minimization process of deformable curves, wherein initially unit-length curves are placed at the discontinuity points and then are attracted to paths of minimum change in the Principal Directions (around a fixed neighbourhood) until they reach other discontinuities. After this stage, the set of final curves will indicate the regions on the object's surface to be represented as separate parts. We propose that globally deformable superquadrics can capture this notion of an object's parts well. Therefore, in our solution we fit deformable superquadrics to each of the segmented regions to get a final extracted model of the object suited for identification of its overall shape and position. Previous results on the segmentation method were reported in [Borges–Fisher (1993)].

Figure 2 shows a functional diagram of the segmentation module of the approach.

3.1 Geon Classification

Using the superquadric parameters for 3D object recognition directly would not be very efficient because they lack expressiveness for indexing into the object database. This problem can be solved by mapping the superquadric parameters into a set of distinctive volumetric shapes with good potential for indexing. In the end, both qualitative and quantitative shape information would be readily available.

One set of distinctive volumetric shapes are the **geons**, which were proposed as part of the Recognition-by-Components theory of Biederman [Biederman (1987)].

The main purposes of the geon classification stage for the recognition approach developed in this work are twofold: first, it is to provide a symbolic index for the volumetric features of the objects' parts with which a more efficient search could be designed; and second, it is to accommodate the range of numeric values, on which the superquadric shape parameters vary, on prototypical volumes that are representative of that class of shapes.

In order to perform the superquadrics into geons classification we developed a Radial Basis Function neural net classifier algorithm, which achieved higher classification rates than other previously published algorithms for the problem at hand, making it possible in our case to deduce more accurate qualitative information for visual matching of three dimensional shapes. The solution was based on a regularization framework, where a especially designed Cross Validation criterion was constructed for achieving convergence.

4 Matching

The matching approach developed in this work is feature based, and it receives its input data from the segmentation and classification modules. Both quantitative and qualitative information about the data are explored: first by means of an indexing of the model library, based on the qualitative features; and second by a constrained search model using the quantitative features of the superellipsoid representation. The constrained search algorithm presents a novel measure of qualitative similarity, which allows ranking of the surviving hypotheses in a way suitable for evaluating class-based recognition of complex 3D objects.

Figure 3 shows how the model library is organized. The indexing keys (G01, G02, G03, G04, G05, G06, G07, G08, G09, G10, G11, G12) represent the twelve geons, which are the symbolic features for the parts of the models. The models containing a specific part are linked to the indexing key for that part.

For a matching process to be able to deal with objects with similar but not identical geometric properties, and with features that may have a small range of values, a mechanism to rank or to measure the quality of the matching is more appropriate than usual binary (*i.e.* matches / fails) measures. In this case inexact matchings have to be considered, and often a categorization of the valid hypotheses is also desirable. Figure 4 shows the stages of the matching approach.

When a hypothesis is generated it is based on evidence extracted from the image and from the knowledge built into the system. A binary evaluation of this hypothesis, being either true or false only, overshadows the fact that because of noise, occlusion, failures or deviations in

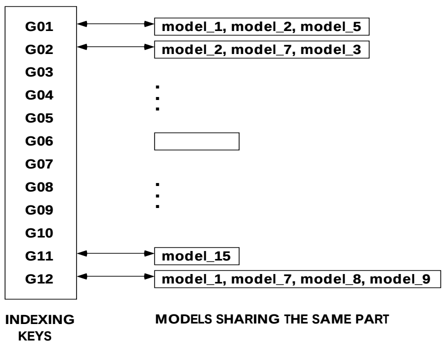


Figure 3: Schematic figure of the model library structure showing the geon classes as indexing keys and hypothetical models that share the pointed geons as parts.

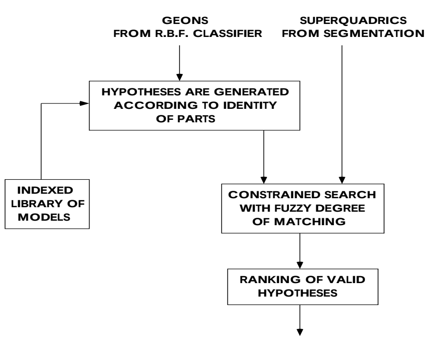


Figure 4: Functional diagram of the matching approach.

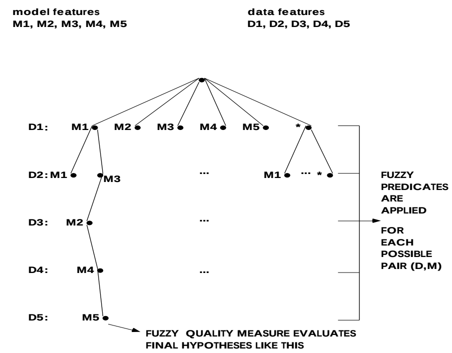


Figure 5: Interpretation tree search model showing where the evaluation of the fuzzy predicates and the fuzzy degree of similarity apply.

the segmentation, and geometric similarity rather than identity between the objects, a match is never exact in a real situation. A quality measure of similarity in the matching process provides a more robust and realistic evaluation of the hypotheses generated, since it can order the hypotheses within the range of 0 - 100% true. Fast information integration capabilities are also required in deriving this measure. We derived such a measure from Possibility Theory, or Theory of Fuzzy Sets [Zadeh (1978)]. Figure 5 shows an example of how the fuzzy predicates are evaluated as part of an Interpretation Tree search.

The matching algorithm, including the evaluations of the fuzzy predicates and the final degree of similarity, is divided into the following steps:

1. Compute normalized distances between the pair of corresponding features while analyzing the current binding (M_i, D_j), using:

$$\delta(fm_q fd_q) = \|1.0 - (fm_q - fd_q)/fm_q\| \qquad 1 \leq q \leq n \qquad (1)$$

   where n is the number of features involved amongst all predicates, in this case $n = 11$ (7 predicates, but the third unary predicate provides actually 5 feature evaluations).

2. Evaluate the degree of membership of each distance feature $\Delta\mathcal{F}$ producing fuzzy inputs. The membership functions are defined in Figure 6. For each feature evaluation computed using $\delta()$, a degree of membership μ (between 0.0 and 1.0) and a linguistic label $\mathcal{L}$ = (VERY FALSE, FALSE, ACCEPTABLY TRUE, TRUE, VERY TRUE) are associated to that feature matching. This can be written as the following fuzzy relationship $R_1 : \Delta\mathcal{F} \times \mathcal{L}$,

$$\mu R_1([\delta f_1, \delta f_2, ..., \delta f_{11}], [l_1, l_2, ..., l_5]) =$$

$$minimum_i[\mu_{lf}(\delta f_i)] \qquad (2)$$

where $[\delta f_1, \delta f_2, ..., \delta f_{11}] \in \Delta\mathcal{F}$,
and $[l_1, l_2, ..., l_5] \in \mathcal{L}$.

3. The actual fuzzy predicates are thresholded to indicate acceptable matchings (the ones that satisfy the constraints). The acceptable values are ACCEPTABLY TRUE, TRUE, and VERY TRUE.

4. Accepted pairings are given a fuzzy degree of similarity, which is computed as follows. Weights are given according to the values shown in Table 1 to express the relevance of each feature matched and its degree of membership. The values were determined empirically, reflecting our experience with the tests on the importance of each feature. A membership function $R_2 : \mathcal{L} \times Sim\_set$ is defined as:

$$\mu R_2[l_{f1}, l_{f2}, ..., l_{f11}, Sim\_set] = \frac{[\sum_{i=1}^{11} wf[i] \times wt[j]]}{maximum_j(\sum_{i=1}^{11} wf[i] \times wt[j])} \qquad (3)$$

where, $j \in [1, 2, 3, 4, 5]$ (*i.e.* 5 fuzzy sets), and Sim_set is chosen to be the same as in Figure 6 (b).

5. The quality measure of similarity is computed by combining the fuzzy relationships R_1 and R_2 by a composition rule of Fuzzy Sets:

$$\mu R_1 \circ R_2([\delta f_1, \delta f_2, ..., \delta f_{11}], Sim\_set) = max_l \;\; min[\mu R_1([\delta f], [l]), \mu R_2(l_f, Sim\_set)] \qquad (4)$$

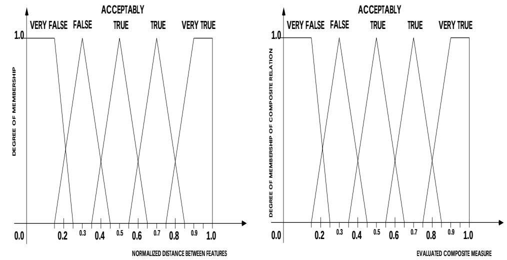
6. Steps 1 to 5 are repeated until all the data parts are evaluated. The surviving hypotheses are then ranked as follows. The tree evaluation produces np fuzzy quality measures, one for each part, and a global similarity measure is computed by evaluating a membership function $R_3 : \mathcal{L} \times Sim\_set$, defined as:

$$\mu R_3[l_{part_1}, ..., l_{part_{np}}, Sim\_set] = \frac{[\sum_{i=1}^{np} wpart[i] \times wt[j]]}{maximum_j(\sum_{i=1}^{np} wpart[i] \times wt[j])} \qquad (5)$$

with Sim_set as in Figure 6 (b), and the weights for the parts $wpart[i]$ are set as equal to 1 for all the np parts.

5 Pose Estimation

Pose estimation is a necessary step to locate the successfully matched hypotheses. Pose estimation of 3D constrained objects is a difficult and important problem for 3D model based object recognition. The method presented here is



(a) (b)

Figure 6: (a) Membership functions; (b) Similarity Sets.

$\mathbf{wf_1}$	$\mathbf{wf_2}$	$\mathbf{wf_3}$	$\mathbf{wf_4}$	$\mathbf{wf_5}$	$\mathbf{wf_6}$
2.0	1.0	1.0	1.0	1.0	1.0
$\mathbf{wf_7}$	$\mathbf{wf_8}$	$\mathbf{wf_9}$	$\mathbf{wf_{10}}$	$\mathbf{wf_{11}}$	
1.0	2.0	3.0	3.0	3.0	—

$\mathbf{wt_1}$	$\mathbf{wt_2}$	$\mathbf{wt_3}$	$\mathbf{wt_4}$	$\mathbf{wt_5}$
2.0	1.5	1.0	0.5	0.2

Table 1: Weights expressing the relevance of each feature (wf_i) and each fuzzy set (linguistic term) (wt_j) in the similarity evaluation.

based on the Kalman Filtering framework, which is a general technique for state and parameter estimation of stochastic systems.

A parameterized or constrained model of a 3D object is a description which imposes general constraints between the model components. These constraints can define equality relations such as: co-linearity, co-planarity, angle relationships, and edge length relationships; or they can define inequality relations such as range of distances or angles. These are typical parameterizations on positions, other types of parameterizations can also involve shape functions.

An articulated object is a particular type of parameterized object having constraints on the spatial locations of the parts.

The method developed in this work deals with pose estimation of 3D constrained volumetric models. The models are represented by assemblies of superquadrics. The method is based around an Iterated Extended Kalman Filter, which is used as the estimation tool where the problem is formulated. The technique is applied as a two stage approach. First, the pose of each part of the model is estimated using characteristic points from the superquadric representation. In the second stage these initial estimates are considered as the *a priori* estimates of an IEKF, the constraints on the model are appended as a single measurement with zero uncertainty, and with the constraint equations linearized they are fused into an IEKF in a batch mode.

The method is applied to estimate the pose of the valid hypotheses found in the matching stage, and results are shown for nine different test scenes.

6 Experiments and Results

This section presents results on matching complex 3D objects from range data against a model library of 16 objects. The model library is indexed by the qualitative volumetric parts of the objects.

Range images acquired were processed by the segmentation and classification procedures. The direct input to the matching procedure is an array of measurements with the quantitative features of the superquadrics (15 features), and the qualitative feature identifying the geon category for each volumetric part successfully segmented in the range image. After the hypotheses are formed and ranked a verification is made by computing the poses of the articulated objects using an algorithm based on a Kalman Filtering framework.

Results are shown for six different objects: a wooden "doll", and plastic miniatures of two "horses", a "cow", a "giraffe", and a "kangaroo". The second object "horse" was scanned in four different viewing positions for testing the recognition approach. Figure 7 shows the original range images of the objects, and Figure 8 shows the results with the matched models overlayed on the original 3D data.

7 Conclusions

We have presented here a solution to the problem of recognizing 3D complex curved objects using parameterized volumetric models. Further results and details of the solution are in [Borges (1995)]. The approach assumes objects can be represented by concave volumetric parts, and many natural shapes and man-made objects can be modelled by such primitives [Marr–Nishihara (1978)], [Pentland (1986)]. Recent studies in Psychology suggest that parts structuring plays an important role in object recognition processes [Biederman (1987)].

The domain of the geometric shapes explored is that of complex curved objects with articulated parts, where there is a great deal of similarity between some of the parts. These objects are exemplified by animal shapes, however the general characteristics and complexity of these shapes are present in a wide range of other natural and man-made 3D objects.

Most of the previous work in recognition has dealt with objects that can be approximated by few planar surface patches. Few works have used higher level representation primitives in order to describe complex curved 3D objects, and to capture the notion of parts structure in depicting the forms. However, these attempts have addressed different, and not all, elements of the model-based 3D object recognition problem, namely parts-based segmentation and description [Bajcsy–Solina (1987)], [Pentland (1986)], [Gupta (1991)], and generic symbolic identification without location [Nevatia–Binford (1977)], [Marr–Nishihara (1978)], [Dickinson et al. (1992)], for a restricted set of forms and with no allowance for articulated parts.

This paper presented a novel approach to the 3D recognition problem. Major contributions presented in this work include:

- A novel approach to recognizing complex 3D objects using parameterized parts-based models. It addresses the issues of representation, feature extraction, and interpretation providing answers to each of them by exploiting the scenario of objects exemplified by animal shapes.

- A segmentation into parts method that integrates a variety of techniques to provide a solution to parts-based segmentation of 3D objects. The approach has three basic stages: first, it computes stable and accurate estimates of the differential structure of the object's surfaces; second, it finds closed boundaries of regions in the data by linking detected surface discontinuities (depth, concave orientation, and negative minima of curvature) using distributed active contours; and third, it fits deformable superquadric models to each of the closed regions.

- A shape representation which features both quantitative and qualitative information is introduced. It has a large scope of applicability, and indexing properties to represent natural articulated 3D forms.

- A multi-dimensional classifier technique for mapping deformable superquadric shapes into a subset of geon primitives. The classifier is designed using a Radial Basis Function neural network architecture, and an algorithm to train the network to achieve a high generalization score is presented for the problem.

- An indexing technique, that uses the geon classification of the object parts to organize the model library, and to reduce the size of the search space for matching. This is achieved by focusing the first part of the search on models with equivalent indices.

- A constrained search algorithm for matching parameterized parts-models to the parts-descriptions extracted from the range data. The algorithm builds an Interpretation Tree with the ordered hypotheses and features, and decides the pairings by evaluating fuzzy predicates using a pre-defined decision space. This represents precisely ranges of distinctness in the feature values, and attributes different levels of confidence. These local fuzzy measures in the pairings are further used in a global quality measure of the match.

- A similarity metric to evaluate matches. A pre-defined set of weights is given for the feature predicates and for the fuzzy measures. The quality measure of the match describes the degree of membership of the weighted set of variables relative to what would be a perfect match.

- A pose estimation algorithm for articulated objects represented by parameterized volumetric parts. The poses are computed in order to locate and verify the best hypotheses found in the matching stage of recognition. A method for pose estimation of individual deformable superquadric parts is shown, and is used as the initial estimates for computing the pose of articulated objects in a second stage. Both stages are developed in a Kalman Filtering framework. The algorithm shows how to include constraints in the estimation process, and it achieves a final pose integrating all the information.

Acknowledgements

This research is partially supported by CNPq, through grants 200389/88.2 and now 520441/95-8. This work contributed from discussions with R. Fisher, A. Fitzgibbon, M. Trucco, and D. Eggert from DAI, U. Edinburgh.

8 References

R. Bajcsy, F. Solina, Three Dimensional Object Representation Revisited. *Proc. First Int. Conf. Computer Vision*, London, IEEE Press, (1987) 231–240.

I. Biederman, Recognition-by-Components: A theory of human image understanding. *Psychological Review* **94** (1987) 115–147.

D. L. Borges, *Recognizing Three-Dimensional Objects using Parameterized Volumetric Models*, PhD Thesis, Dept. Artificial Intelligence, University of Edinburgh, 1995.

D. L. Borges, R. Fisher, Segmentation of 3D Articulated Objects by Dynamic Grouping of Discontinuites. *Proc. British Machine Vision Conference 93*, Guildford, BMVA Press, (1993) 279–288.

S. Dickinson, A. Pentland, A. Rosenfeld, From Volumes to Views: An Approach to 3-D Object Recognition. *CVGIP: Image Understanding*, **55(2)** (1992) 130–154.

R. Fisher, *From Surfaces to Objects: Computer Vision and Three Dimensional Analysis*, John Willey, 1989.

P. Flynn, A. Jain, BONSAI: 3-D Object Recognition Using Constrained Search. *IEEE Trans. Pattern Analysis and Machine Intelligence* **13(10)** (1991) 1066–1074.

W. E. L. Grimson, On the recognition of parameterized 2D objects. *Int. Journal of Computer Vision* **2** (1988) 353–372.

A. Gupta, *Surface and Volumetric Segmentation of Complex 3-D Objects Using Parametric Shape Models*, PhD Thesis, Dept. Computer Science, University of Pennsylvania (USA), 1991.

A. Jain, R. Hoffman, Evidence-Based Recognition of 3-D Objects. *IEEE Trans. Pattern Analysis and Machine Intelligence* **10(6)** (1988) 783–802.

D. Marr, K. Nishihara, Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. London B* **200** (1978) 269–294.

R. Nevatia, T. Binford, Description and Recognition of Curved Objects. *Artificial Intelligence* **8** (1977) 77–98.

A. Pentland, Perceptual Organization and the Representation of Natural Form. *Artificial Intelligence* **28** (1986) 293–331.

L. Zadeh, Fuzzy Sets as a Basis for a Theory of Possibility. *Fuzzy Sets and Systems* **1(1)** (1978) 3–28.

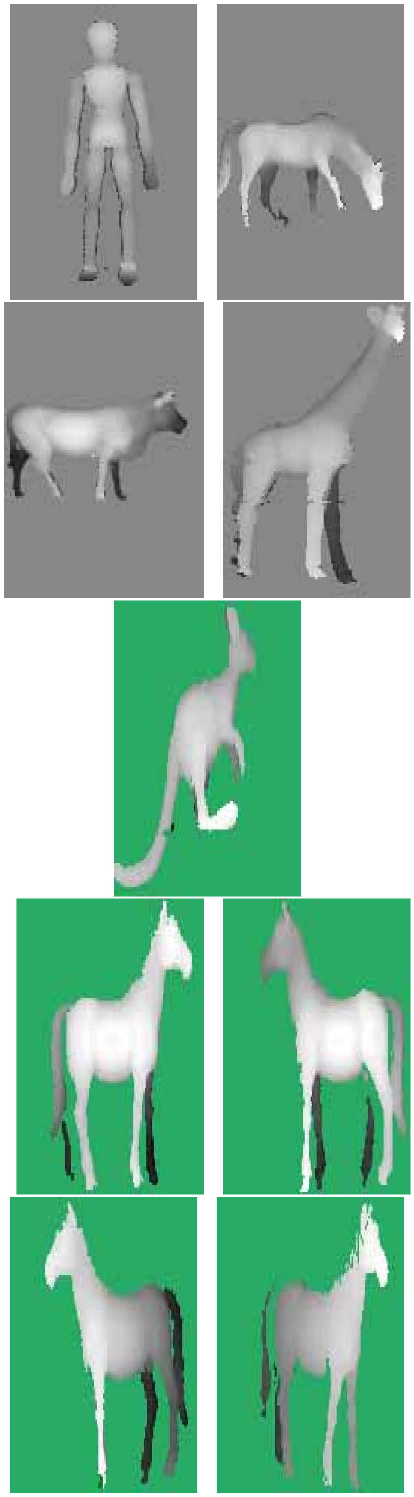


Figure 7: Range images of objects tested for recognition. (a) “doll”; (b) “horse (first)”; (c) “cow”; (d) “giraffe”; (e) “kangaroo”; (f)(g)(h)(i) “horse (in four different viewing positions)”.

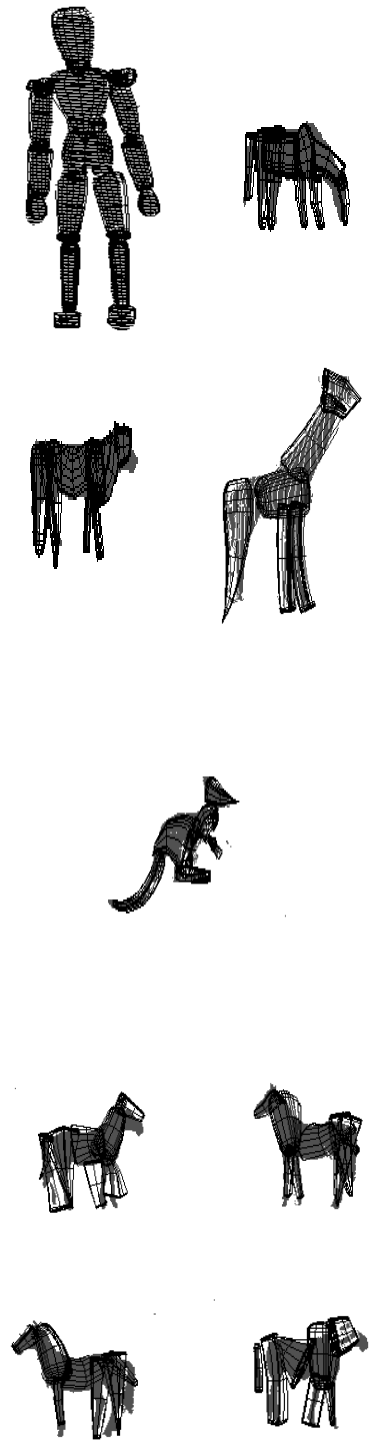


Figure 8: Recognized objects with models overlayed on 3D data in final estimated position. (a) “doll”; (b) “horse (first)”; (c) “cow”; (d) “giraffe”; (e) “kangaroo”; (f)(g)(h)(i) “horse (in four different viewing positions)”.