Class-based recognition of 3D Objects represented by volumetric primitives

Díbio L. Borges and Robert B. Fisher
Dept. of Artificial Intelligence, University of Edinburgh
5 Forrest Hill, Edinburgh EH1 2QL (United Kingdom)
{dibio, rbf}@aifh.ed.ac.uk

Abstract

This paper presents a novel approach to recognizing 3D complex objects that have similar geometric structure but belong to different subclasses. Test scenes are acquired by a laser striper as range images, and the objects are modelled using a composite volumetric representation of superquadrics and geons. Matching is decomposed into two stages: first, an indexing scheme designed to make effective use of the symbolic keys of the representation is implemented in order to direct the search to the models containing the parts identified; 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.

1 Introduction

This paper is concerned with a matching approach developed for model-based recognition. The approach is feature based, and it receives its input data from a segmentation [3] and classification modules (see Figure 1 (a)). After matching, localization and verification are performed in a complete recognition process [2]. This paper focuses on the matching process developed for recognition. 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 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.

Two important issues in the design of the matching process for recognition are efficiency and robustness. The first is concerned with finding the matches as quickly as possible, avoiding misleading paths sooner rather than later, and the second is concerned with error or uncertainty management where false and inexact matchings are to be dealt with. One way to address efficiency is by splitting the search optimization problem into two related ones: 1) minimization of the search space and 2) minimization of the portion to be explored in the search. In our approach we use volumetric primitives (deformable superquadrics and geons) to represent the objects and this has the effect of reducing drastically the search space, since the number of features is directly related to the size of the search. To minimize the fraction of the search space explored in the matching process, one general strategy would be to group in a meaningful way the objects in the database, so common characteristics shared by subsets of the objects would be first identified and then trigger the search in subsets of the database.

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 1 (b) shows a functional diagram of the matching process developed in this work in order to achieve recognition. The approach is structured to favour efficiency and robustness, and works particularly well in the domain of this work.

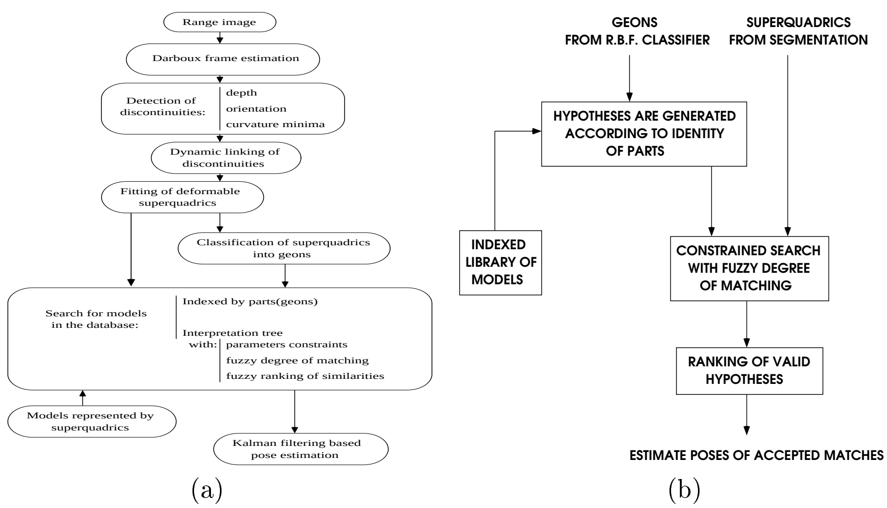


Figure 1: (a) Functional diagram of the complete recognition approach; (b) The matching approach presented here.

2 Parametric Forms and Symbolic Features

A problem in 3D model-based recognition is to find a rich and suitable representation for describing the classes of objects which the system aims to recognize. In our opinion there is no general representation that can describe efficiently all types of shapes. Having said that, there are criteria[7] upon which we could judge existing and newly proposed representations for 3D recognition. In general, 3D volumetric primitives perform better than other commonly used representations such as surfaces, contours, and points in practical design issues like primitive complexity, model complexity, search complexity, reliance on verification, and model flexibility. To describe a wide range of complex articulated 3D shapes for the pur-

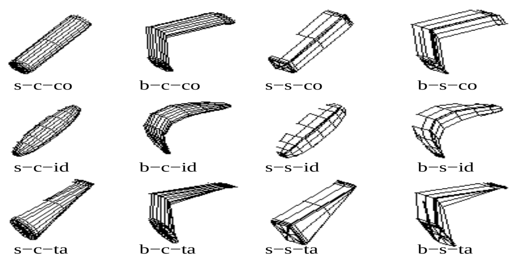


Figure 2: The set of twelve (12) geons modelled using the attributes of axis shape, cross-section edge, and cross-section size.

pose of recognition we use two volumetric representations which complement each other's properties: superquadrics [8] and geons[1]. Superquadrics are attractive for use in Computer Vision mainly because of their representational power, and the small number of parameters necessary to define them. However, it has been argued [6] that one of their drawbacks is the difficulty in achieving shape discrimination, or the lack of good indexing properties by their parameters. To address this problem we derive a classification of volumetric shapes (a subset of geons, see Figure 2) and form a composite representation, superquadrics and geons. This results in a powerful representation for object recognition since we will have the representational capacity of the superquadrics and the indexing basis provided by the geons. All the models used in this work were created manually using the THINGWORLD [8] modelling system, to shape and position the superquadrics, and then their descriptions were edited and placed in the database of models.

The geometric models are hierarchical assemblies of volumetric features. The model primitives are superquadrics, parameterized and described by: ϵ_1, ϵ_2 parameters of the superquadric; a_x, a_y, a_z, superquadric axis lengths, t_x, t_y x,y tapering deformation; and *rad_bend* radius of bending deformation.

A model structure hierarchy is formed by placing the primitives relative to each other, relative to a global coordinate system, specifying the transformation by X, Y, and Z for translation, and R, S, and T for the rotation components. Movable joints are specified by labelling a part allowed to move, and a distance limit (measured as an euclidean distance between the mass centres of the connected parts) which specifies the maximum relative movement. Additional information in the model is primitive adjacency, indicating which parts are adjacent to each other, and labels for the geon classes the primitives belong to.

3 Finding and Ranking Correspondences

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, ignores the fact that, because of noise, occlusion, failures or deviations in the segmentation, and geometric similarity rather than identity between the objects, a match is never exact in a real situation. A qualitative 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 derive such a measure from Possibility Theory, or Theory of Fuzzy Sets [9].

One suitable way to organize the search is to pose the problem as a constrained search of a tree of interpretations, known as the Interpretation Tree approach. The Interpretation Tree approach can be seen as an instance of the hypothesize-and-test paradigm in Artificial Intelligence. At each level down the tree, going from the root to the leaf nodes, a different feature is tested against the model, thus building up the feasible matchings between data and model. Other researchers have used the Interpretation Tree in Computer Vision, most notably [5, 4].

A verification stage including the pose estimation of the model is done after valid interpretations of the data are found and ranked. A variation of the Interpretation Tree approach estimates the transformation (rotation and translation) between data and model while building the interpretation [5, 4]. This is a powerful constraint, in particular when using features like edges and surfaces which generate thousands of possible bindings to be searched. Because we use volumetric primitives, and we use an indexing scheme to organize the models and narrow down the initial hypotheses, we deal with a small number of possible interpretations (tens or hundreds). The surviving hypotheses are therefore ranked by the fuzzy evaluation process and passed to the pose estimation module which introduces a new method to calculate poses of articulated objects represented by volumetric primitives. The approach presented here is efficient and it is specially suited for our problem of recognizing similar complex objects.

3.1 Geometric Constraints and Design Features

The efficacy of the Interpretation Tree approach as a recognition algorithm is based on the use of geometric constraints to prune the branches of the tree which lead to invalid hypotheses. These constraints are forced by unary and binary predicates, which are applied sequentially to every hypothesis generated as the matching grows by adding new bindings. If a predicate is not satisfied the interpretation is discarded, and the subtree is pruned.

Unlike the usual Interpretation Tree, which uses a bivalued predicate for evaluating the geometric constraints in the search, we developed a Fuzzy Predicate with five (5) fuzzy sets indicating possible evaluations of the evidence. This allows us to deal with uncertainty in the matching in a more robust way, and also to have a ranking of surviving hypotheses, which proves to be a distinctive measure in matching objects with similar geometric properties.

In the remainder of this section we describe the unary and binary fuzzy predicates, and the special design features of our Interpretation Tree algorithm respectively. The details of the fuzzy predicate evaluation, together with the quality measure of similarity are given in Section 3.2.

In our approach we use four different fuzzy unary predicates:

1. **The Volume Predicate** - comparing (using $\delta()$) the estimated volume and the maximum volume prescribed in the model.

2. **The Geon Type Predicate** - comparing (using $\delta()$) the similarity between the modelled and extracted types.

3. **The Shape Intrinsic Parameter Predicate** - compares (using $\delta()$) the five shape parameters of the deformable superquadric that identify the geon $(\epsilon_1, \epsilon_2, t_x, t_y, rad\_bend)$.

4. **The Part Scale Predicate** - compares (using $\delta()$) the ratio between the longest axis, and the shortest axis (*i.e.* from the a_1, a_2, a_3) of the two volumetric primitives.

We use three binary predicates:

1. **The Adjacency Predicate** - evaluates (using $\delta()$) feature adjacency for each model and data feature pairing.

2. **The Distance Predicate** - compares (using $\delta()$) the Euclidean distance between each pair of associations made to see if they agree within a specified tolerance. The distance is measured between the mass centres of each volumetric part. In the case where the part is articulated the tolerance given is higher in order to comply with the allowed range of distance between the two parts.

3. **The Parts Proportion Predicate** - evaluates (using $\delta()$) the proportion between the biggest and smallest parts of the model and data to provide discriminative information in the case of similar objects.

In order to allow for spurious data, perhaps from bad segmentation or fitting, we provide the Interpretation Tree with NULL associations, which gives the capacity, during the search, to ignore any data primitive $\mathcal{D}_i$ not found in the model. Data primitives $\mathcal{D}_i$ are ordered starting from the biggest parts until the smallest before being passed to be searched. This improves the efficiency of the search.

3.2 Evaluating Similarity in a Match by a Fuzzy Measure

Feature values extracted from image data, such as parameters of the deformable superquadrics fitted in the segmentation procedure, are evidence for a matching procedure and as such there are variations in the degree of uncertainty for these features. When evaluating a geometric constraint for (data primitive, model primitive) binding, usually a tolerance is prescribed for each predicate test and the matching result indicates only the acceptance or not (*i.e.* true or false) of that binding. In this work we aim to recognize complex 3D objects that are similar in their geometric properties and are organized into classes. This brings to the match evaluation the task of measuring different degrees of similarity, hence the need for a quality measure of similarity.

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:

$$\delta(fm_q, fd_q) = \|1.0 - (fm_q - fd_q)/fm_q\| \quad 1 \le q \le n \tag{1}$$

   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 3 (a). 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 $R_1 : \Delta\mathcal{F} \times \mathcal{L}$,

$$\mu R_1([\delta f_1, \delta f_2, ..., \delta f_{11}], [l_1, l_2, ..., l_5]) = minimum_i[\mu_{lf}(\delta f_i)] \tag{2}$$

   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:

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

   where, $j \in [1, 2, 3, 4, 5]$ (*i.e.* 5 fuzzy sets), and Sim_set is chosen to be the same set as in Figure 3 (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 [9]:

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

wf$_1$	**wf$_2$**	**wf$_3$**	**wf$_4$**	**wf$_5$**	**wf$_6$**	**wf$_7$**	**wf$_8$**	**wf$_9$**	**wf$_{10}$**	**wf$_{11}$**
2.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0	3.0	3.0	3.0

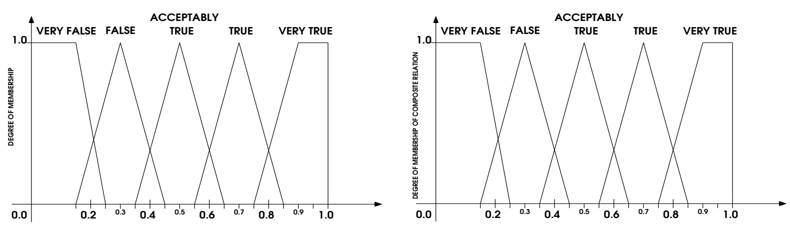
wt$_1$	**wt$_2$**	**wt$_3$**	**wt$_4$**	**wt$_5$**
2.0	1.5	1.0	0.5	0.2

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

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:

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

with Sim_set as in Figure 3 (b), and the weights for the parts $wpart[i]$ are set as equal to 1 for all the np parts.



(a) (b)

Figure 3: (a) Membership functions; (b) Similarity Sets.

4 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 geon parts of the objects. 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. 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 stability. Figure 4 shows the original range images of the objects, and Figure 5 shows the results with the matched models overlayed on the original 3D data.

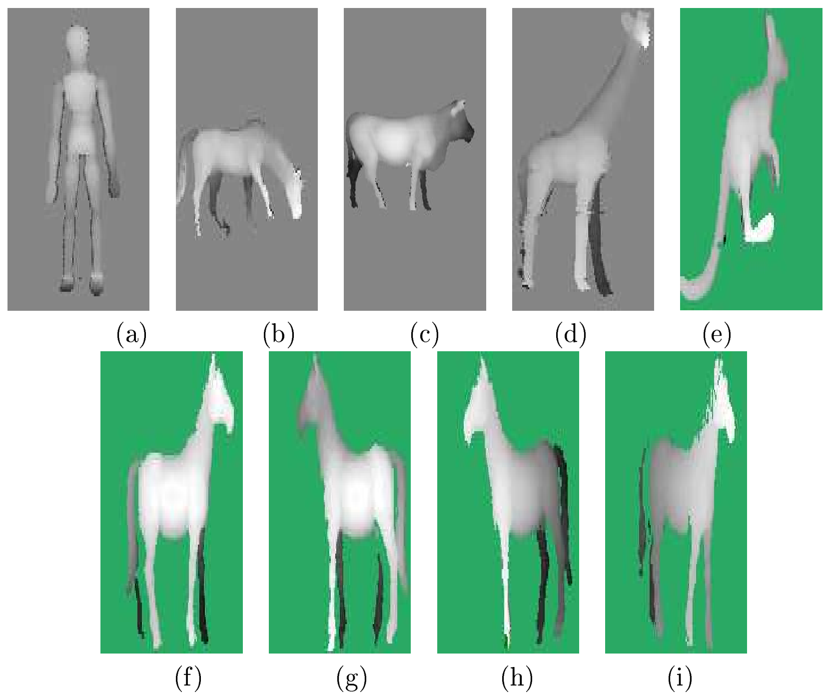


Figure 4: Range images of objects tested for recognition. (a) “doll”; (b) “horse (first)”; (c) “cow”; (d) “giraffe”; (e) “kangaroo”; (f)(g)(h)(i) “horse (in four different viewing positions)”.

Table 2 shows that all objects were successfully recognized and ranked differently according to the fuzzy measure of similarity. Of particular significance are the facts that: 1) most of the objects tested have effectively identical quadruped structure yet the correct models are identified, and 2) the other models ranked for each image are those models that are closest to the data in terms of shape parameters. Thus, we can see good performance on class and subclass identification. Recognition was unsuccessful in four other images of another quadruped – a pig – largely because the leg parts were too small to be extracted reliably [2]. The data primitives ($\mathbf{D}_n$) are ordered according to size, from the largest part to the smallest, and the model primitives are ordered in the sequence the parts are described in the model, when passed to Interpretation. Preliminary pairings are chosen by applying the Geon Type Predicate, and pruning occurs with the evaluation of all the unary and binary fuzzy predicates. The search complexity for the correct model is typically on the order of 100-200 nodes investigated.

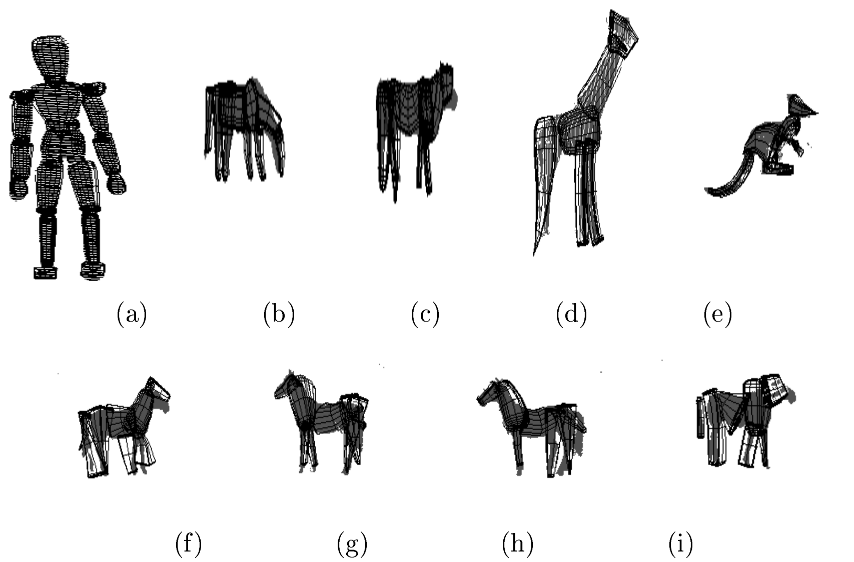


Figure 5: Recognized objects with models overlayed on 3D data in final estimated position. (a) "doll"; (b) "horse (first)"; (c) "cow"; (d) "giraffe"; (e) "kangaroo"; (f)(g)(h)(i) "horse (in four different viewing positions)".

5 Conclusions

This paper has presented an efficient and robust method to perform matching of 3D objects from range image data. 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. The matching technique presented in this paper has many different features addressing issues of part-based representation, inexact matching, library indexing, and constraint based search. Of particular significance is the fact that the recognition process was capable of both recognizing and discriminating between models with the same structure but different shapes (as well as ranking the candidate models in a sensible order relative to the true shape.)

The method divides the problem into two stages: an indexing stage and a constrained search stage with fuzzy degrees of similarity. The indexing stage selects a subset of the model library to be fully searched based on the output of the geon classifier. The models are indexed in the library by grouping them on the common part primitives (geons) they share. This organization allows for robustness in recognition since a meaningful and reliable classification relates these parts to their parameters. This indexing scheme has also the property of sublinear growth with the size of the library, and so it does not compromise its performance when new models are added in the database. A quality measure of similarity is introduced in a constrained interpretation tree based search algorithm providing a robust

Object	Model				F.M.S.				Label			
doll	m1	–	–	–	0.89	–	–	–	V.T.	–	–	–
horse (first)	m7	m2	m14	–	0.82	0.61	0.54	–	T.	T.	A.T.	–
cow	m2	m7	m14	m13	0.92	0.61	0.48	0.41	T.	T.	A.T.	A.T.
giraffe	m10	m6	–	–	0.94	0.41	–	–	V.T.	A.T.	–	–
kangaroo	m16	–	–	–	0.74	–	–	–	T.	–	–	–
horse (v.1)	m7	m2	m14	–	0.58	0.50	0.45	–	A.T.	A.T.	A.T.	–
horse (v.2)	m7	m2	m14	–	0.70	0.62	0.58	–	T.	A.T.	A.T.	–
horse (v.3)	m7	m2	m14	–	0.68	0.61	0.52	–	T.	A.T.	A.T.	–
horse (v.4)	m7	m2	m14	–	0.50	0.42	0.40	–	A.T.	A.T.	A.T.	–

Table 2: Results showing the matched objects. All hypotheses higher than ACCEPTABLY TRUE (A.T.) are shown. T. = TRUE, V.T. = VERY TRUE, and F.M.S. = Fuzzy Measure of Similarity. Models mentioned are: m1(doll), m2(cow), m6(nessie), m7(horse), m10(giraffe), m13(hippo), m14(bear), and m16(kangaroo).

and efficient mechanism to rank and compare inexact matchings. The quality measure is built by using the Theory of Fuzzy Sets [9], first by evaluating feature bindings (D, M) (Data primitives, Model primitives) with fuzzy predicates, and deriving a fuzzy degree of similarity which ranks the valid hypotheses reached in the constrained search. Results are given for a variety of complex 3D shapes, and identification and ranking are successfully achieved.

References

[1] I. Biederman. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94:115–147, 1987.

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

[3] D.L. Borges and R.B. Fisher. Segmentation of 3D articulated objects by dynamic grouping of discontinuities. In J. Illingworth, editor, *Proc. British Machine Vision Conf. 93, Guildford*, pages 279–288. BMVA Press, 1993.

[4] O.D. Faugeras and M. Hebert. The representation, recognition, and locating of 3-D objects. *I.J. Robotics Res.*, 5(3):27–52, 1986.

[5] W.E.L. Grimson and T. Lozano-Perez. Model-based recognition and localization from sparse range or tactile data. *I.J. Robotics Res.*, 3(3):3–35, 1984.

[6] R. Jain and A.K. Jain. Report on range image understanding workshop, east lansing, michigan, march 88. *Machine Vision and Applications*, 2:45–60, 1989.

[7] D. Marr. *Vision*. W.H. Freeman, 1982.

[8] A. Pentland and S. Sclaroff. Closed-form solutions for physically based shape modeling and recognition. *IEEE Trans. P.A.M.I.*, 13(7):715–729, 1991.

[9] L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1(1):3–28, 1978.