3D Object Recognition Using Shape Similiarity-Based Aspect Graph

Christopher M. Cyr
Div. of Engineering
Brown University
Providence RI 02912
cmc@lems.brown.edu

Benjamin B. Kimia
Div. of Engineering
Brown University
Providence RI 02912
kimia@lems.brown.edu

Abstract

We present an aspect-graph approach to 3D object recognition where the definition of an aspect is motivated by its role in the subsequent recognition step. Specifically, we measure the similarity between two views by a 2D shape metric of similarity measuring the distance between the projected, segmented shapes of the 3D object. This endows the viewing sphere with a metric which is used to group similar views into aspects, and to represent each aspect by a prototype. The same shape similarity metric is then used to rate the similarity between unknown views of unknown objects and stored prototypes to identify the object and its pose. The performance of this approach on a database of 18 objects each viewed in five degree increments along the ground viewing plane is demonstrated.

1 Introduction

1.1 3D Object Recognition

The human visual system has an uncanny ability to recognize objects from single views, even when presented monocularly under a fixed viewing condition. For example, the identity of most of the 3D models in Figure 1 is immediately clear. The issue of whether 3D object recognition should rely on internal representations that are inherently three-dimensional, *e.g.* those described by GEONs [3], superquadrics [2], deformable solids [17], algebraic surfaces [23], and CAD models, or on collections of two dimensional views, *e.g.*, as those represented by aspect graphs [13], is still being explored. Psychophysical evidence for both views is strong [6, 3] and it is likely that the representation is task dependent, *e.g.*, manipulation vs. recognition. This paper examines a bottom-up, viewer-centered view of 3D object recognition.

There is ample literature on recognition from 2D views. Nayar *et al.* [16] use principal component analysis (PCA) on a set of object views to generate a distance between an unknown view and prototypical views. Cootes and Taylor [5] proposed an active appearance model (AAM) which learns a statistical model by training on a series of 2D images. Ullman and Basri [24] represent each view as a linear combination of prototypical views.

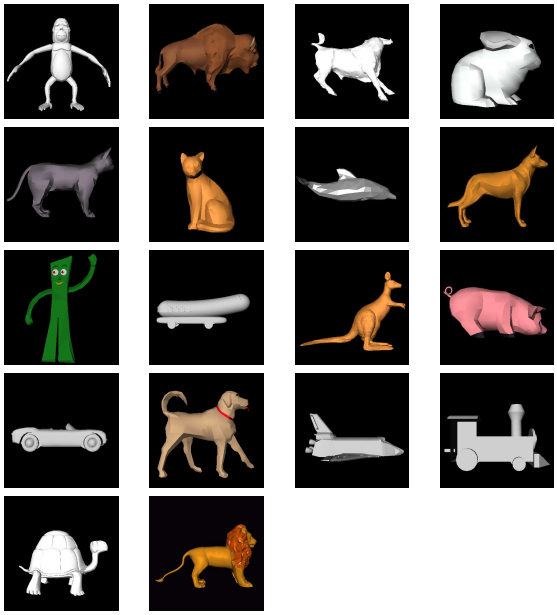

Figure 1. These 3D models represent our initial database of objects which were used to evaluate the performance of the recognition algorithm presented here. (Our database is composed of VRML and Autocad models, taken from www.3dcafe.com)

The *aspect graph* representation [13, 14] identifies regions of the viewing sphere where "equivalent views" and neighborhood relations on the viewing sphere generate a graphical structure of views. Each node of the aspect graph represents a "general view", or aspect, of the 3D object

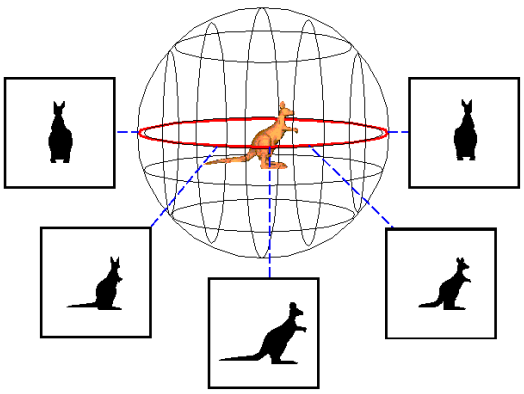

Figure 2. Aspects can be generated by grouping views by the similarity between 2D shapes and representing each aspect by a prototypical view.

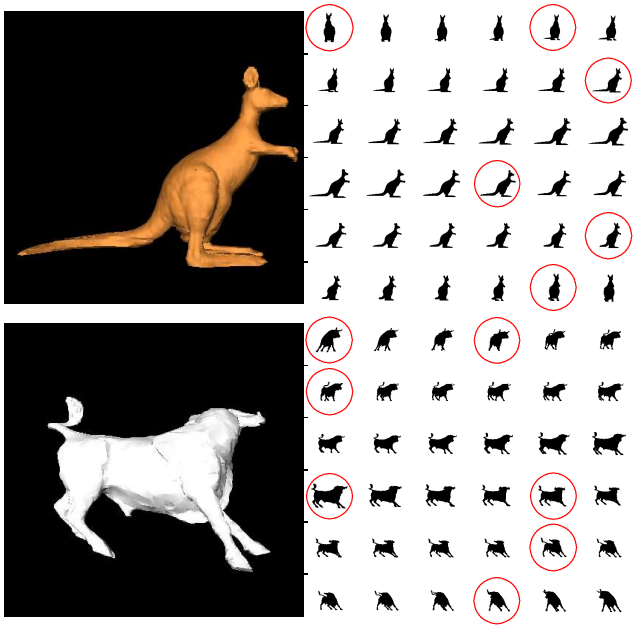

Figure 3. Two objects ("kangaroo" and "bull") from the database are depicted on the left. On the right we show all views from the ground plane in five degree increments. Circled views represent prototypical views.

and represents a maximally connected region on the viewing sphere. Each link represents some visual event where transitions occur between two neighboring general views, namely, the accidental views. Aspect graphs have been defined for polyhedra [22, 21], solids of revolution [9], and curved objects [15]; also, see a survey by Bowyer and Dyer [4]. The main drawback in this type of approach is in the complexity of generating the aspects and in the large number of aspects whose large storage and search requirements are impractical for objects of modest complexity. Eggert *et al.* [10] observed that the aspect graph is often based on a level of detail not fully observable in practice and explored a notion of a scale-space aspect graph to reduce the number of views. Dickinson *et al.* [8] constructed a hierarchical aspect graph system based on a set of primitives. Ikeuchi and Kanade [11] use the similarity of feature extracted from 2D views of the object to form a graph structure used in recognition. The theory which defines the aspect graph, i.e., view stability and view likelihood, was explored by Weinshall and Werman [25].

Our approach to 3D object recognition also relies on a parcellation of the viewing sphere to generate an aspect-graph. Observe that the problem of generating aspects is highly integrated with the problem of recognizing those aspects. The view that aspects are principally generated to reduce the search time from indexing in a database of objects views motivates the use of the recognition scheme itself, rather than the singularities of visual mapping, in defining the aspect. Specifically, the shape similarity metric which is used to judge the distance between two views should also be used to generate aspects: views where similarity varies rather slowly should generate larger size aspects, *e.g.*, side views of four-legged animals, while regions of rapid change in similarity should lead to smaller aspects, *e.g.*, frontal views of four-legged animals.

More specifically, the shape similarity metric between object outlines endows the viewing sphere with a metric which can be used to cluster views into aspects, and to represent each aspect with a prototypical view. In this view, aspect generation is a "segmentation" of viewing sphere and representing each aspect by a prototype. 3D object recognition is then tantamount to indexing a target view into the database of prototype views and ordering these by the shape similarity metric. Two key issues are addressed. First, how can a set of aspects and their prototypical views be generated in order to cope with the requirements of recognition. Second, under what conditions is this set of prototypical views sufficient to achieve robust recognition.

Our earlier work [7] showed that for a single three-dimensional model, the pose of an unknown view can be robustly and accurately determined by a hierarchical comparison of shape similarity of the unknown view against a hierarchically arranged, multiresolutional set of views. Specifically, this work determined the pose of spiral vertebra from single X-ray images by comparison to views generated from a CT-generated 3D model to within 1-2 degrees. We now show that not only pose, but also the identity of an object

can be determined via a 2D shape similarity metric.

We made a number of simplifying assumptions to explore this concept. First, we only considered views generated on one viewing plane, namely, those generated by "walking around" the object, Figure 2. This reduces the number of views for this preliminary stage which is aimed at proving the concept, but is not a fundamentally limiting assumption. Second, we assume that the object is distinct enough from the background that each view can be easily segmented.

Our approach is then summarized as follows. Each 3D object model gives rise to a number of prototypical views, Figure 3, which are selected from the range of possible views for each object in the database, Figure 1, as described in Section 2. We examine two shape similarity metrics, one based on curve matching and the other based on shock graph matching, and conclude that the latter is more suited for generating aspects and for recognition. The recognition of an unknown view is done by matching it with stored prototypes for each object and by ordering them according to the shape similarity metric. The resulting match gives the identity of the object as well as its pose, Section 3. We discuss matching results in Section 4.

2 Similarity-Based Aspects

We now discuss how a shape similarity metric, or alternatively, a distance metric between two shapes can be used to generate aspects on the viewing sphere. Let there be N objects $\{O_1, O_2, ..., O_N\}$, whose silhouettes form M distinct views $\{V_n^1, V_n^2, ..., V_n^M\}$ for each object O_n. Our full database of views then consists of $N \cdot M$ views. Let the distance between two shapes or views V_n^m and V_i^j be denoted by $d(V_n^m, V_i^j)$. Let each aspect be defined by a contiguous collection of views around a prototype V_n^m,

$$A_n^m = \{V_n^{m+l^-}, ..., V_n^m, ..., V_n^{m+l^+}\},$$

i.e., where l^+ and $-l^-$ are defined as the left and right radius of the aspect and which depend on the object n and the prototypical view V_n^m. An aspect is typically defined as the set of generally equivalent views bounded by accidental views. In contrast, we adopt a "region growing" rather than "edge-based" view and use shape similarity to define equivalent views. We must first impose several conditions on the metric. First, the shape similarity metric should be increasing as the viewing angle between two shapes is increased for some local neighborhood:

Criteria 1 (Local Monotonicity) *For each view V_n^m, there exists a $\delta > 0$ such that*

$$d(V_n^m, V_n^{m+i}) < d(V_n^m, V_n^{m+j}) \text{ if } |i| < |j| \leq \delta_0 \quad (1)$$

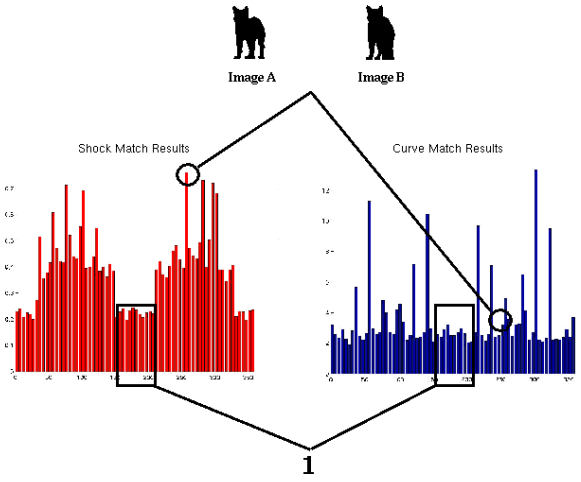

Figure 4. The ideal metric should be sensitive to sudden changes in shape due to changes in view. The shock matching is more sensitive to these transitions as indicated by the sample graph of the derivative of this distance with change in viewing angle for the shock matching metric (left) and the curve matching metric (right). Both indicate large regions of similarity well (1), but curve matching misses some sudden edges (2).

Figure 6 plots the shape similarity metric for a selection of eight views (those are in fact the prototypical views) as a function of differences in viewing angle from each view. Note that the similarity metric we use is monotonically increasing for some neighborhood around each view. Second, each aspect and its prototype should somehow differentiate between aspect and non-aspect views. But, first we discuss our choice of shape similarity metric, which is critical to the success of the recognition system.

2.1 Shape Similarity Metric

In previous work [20, 18, 12, 19], we have defined shape similarity metrics based both on curve matching and on shock graph matching. Both are metrics by construction as they arise from optimal paths of deformation. However, the curve-based metric d_c is generally much faster to compute than the shock-based metric d_s, typically 0.8 minutes vs. 5 minutes when run on an SGI Indigo II, 195 Mhz. Both metrics satisfy criteria 1 in that they both depict local monotonicity with viewpoint variation, although the extent

	-90	-85	-80	-75	-70	-65	-60	-55	-50	-45	-40	-35	-30	-25	-20	-15	-10	-5	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90
-90	–	**0.38**	**0.52**	**0.57**	**0.60**	**0.66**	**0.62**	**0.74**	0.81	0.82	0.80	0.79	0.80	0.79	0.78	0.77	0.81	0.83	0.96	0.82	0.85	0.80	0.80	0.77	0.89	0.79	0.87	0.81	0.77	0.82	0.81	0.70	0.68	0.72	0.72	0.70	0.67
-85	**0.38**	–	**0.57**	**0.61**	**0.67**	**0.68**	**0.68**	**0.79**	0.86	0.87	0.81	0.79	0.79	0.79	0.79	0.79	0.83	0.83	1.00	0.81	0.86	0.81	0.82	0.81	0.93	0.79	0.92	0.85	0.83	0.88	0.81	0.74	0.72	0.71	0.75	0.72	0.72
-80	**0.52**	**0.57**	–	**0.38**	**0.47**	**0.50**	**0.70**	**0.69**	0.80	0.81	0.80	0.80	0.81	0.82	0.80	0.79	0.80	0.83	0.94	0.83	0.81	0.84	0.82	0.75	0.89	0.72	0.81	0.77	0.72	0.85	0.78	0.73	0.72	0.71	0.73	0.70	0.74
-75	**0.57**	**0.61**	**0.38**	–	**0.34**	**0.43**	0.66	0.59	0.79	0.82	0.81	0.78	0.77	0.78	0.78	0.78	0.79	0.85	0.96	0.84	0.85	0.82	0.80	0.70	0.87	0.67	0.87	0.73	0.69	0.86	0.78	0.72	0.67	0.68	0.67	0.68	0.71
-70	**0.60**	**0.67**	**0.47**	**0.34**	–	**0.31**	0.67	0.54	0.73	0.78	0.80	0.78	0.78	0.77	0.79	0.78	0.77	0.80	0.91	0.81	0.81	0.81	0.79	0.68	0.85	0.66	0.81	0.68	0.68	0.84	0.72	0.67	0.67	0.69	0.70	0.68	0.71
-65	**0.66**	**0.68**	**0.50**	**0.43**	**0.31**	–	0.66	0.55	0.66	0.73	0.76	0.80	0.79	0.76	0.79	0.77	0.79	0.77	0.91	0.78	0.77	0.79	0.78	0.67	0.85	0.68	0.80	0.69	0.68	0.86	0.77	0.67	0.67	0.72	0.73	0.75	0.76
-60	0.62	0.68	0.70	0.66	0.67	0.66	–	**0.27**	**0.42**	**0.43**	0.60	0.63	0.64	0.48	0.63	0.64	0.68	0.52	0.62	0.51	0.50	0.65	0.67	0.55	0.70	0.57	0.55	0.57	0.55	0.69	0.68	0.67	0.64	0.70	0.69	0.65	0.71
-55	0.74	0.79	0.69	0.59	0.54	0.55	**0.27**	–	**0.44**	**0.45**	0.59	0.61	0.64	0.53	0.60	0.64	0.69	0.54	0.66	0.56	0.55	0.68	0.68	0.58	0.70	0.61	0.61	0.64	0.63	0.76	0.71	0.72	0.70	0.70	0.71	0.68	0.70
-50	0.81	0.86	0.80	0.79	0.73	0.66	**0.42**	**0.44**	–	**0.24**	**0.25**	**0.30**	**0.31**	**0.36**	**0.38**	**0.38**	**0.39**	0.42	0.58	0.48	0.43	0.43	0.40	0.43	0.41	0.43	0.46	0.47	0.47	0.61	0.57	0.66	0.66	0.68	0.73	0.79	0.80
-45	0.82	0.87	0.81	0.82	0.78	0.73	**0.43**	**0.45**	**0.24**	–	**0.22**	**0.27**	**0.29**	**0.34**	**0.34**	**0.33**	**0.36**	**0.39**	0.58	0.45	0.41	0.39	0.40	0.41	0.43	0.42	0.43	0.47	0.46	0.57	0.57	0.71	0.68	0.73	0.77	0.83	0.84
-40	0.80	0.81	0.80	0.81	0.80	0.76	0.60	0.59	**0.25**	**0.22**	–	**0.25**	**0.28**	**0.29**	**0.32**	**0.31**	**0.35**	0.37	0.53	0.43	0.38	0.35	0.37	0.41	0.40	0.41	0.43	0.48	0.46	0.59	0.58	0.67	0.69	0.67	0.75	0.80	0.78
-35	0.79	0.79	0.80	0.78	0.78	0.80	0.63	0.61	**0.30**	**0.27**	**0.25**	–	**0.26**	**0.31**	**0.30**	**0.29**	**0.30**	**0.36**	0.45	0.39	0.37	0.38	0.38	0.37	0.41	0.41	0.44	0.47	0.43	0.58	0.55	0.66	0.66	0.67	0.72	0.79	0.80
-30	0.80	0.79	0.81	0.77	0.78	0.79	0.64	0.64	**0.31**	**0.29**	**0.28**	**0.26**	–	**0.27**	**0.27**	**0.26**	**0.30**	**0.34**	0.44	0.38	0.35	0.35	0.37	0.36	0.36	0.39	0.42	0.44	0.43	0.57	0.56	0.66	0.66	0.69	0.75	0.79	0.79
-25	0.79	0.79	0.82	0.78	0.77	0.76	0.48	0.53	**0.36**	**0.34**	**0.29**	**0.31**	**0.27**	–	**0.25**	**0.28**	**0.30**	**0.31**	0.50	0.35	0.33	0.31	0.35	0.35	0.37	0.39	0.40	0.46	0.41	0.57	0.57	0.72	0.67	0.71	0.74	0.80	0.80
-20	0.78	0.79	0.80	0.78	0.79	0.79	0.63	0.60	**0.38**	**0.34**	**0.32**	**0.30**	**0.27**	**0.25**	–	**0.25**	**0.28**	**0.26**	**0.44**	0.32	0.30	0.29	0.31	0.35	0.35	0.39	0.40	0.44	0.40	0.53	0.53	0.65	0.66	0.69	0.73	0.78	0.77
-15	0.77	0.79	0.79	0.78	0.78	0.77	0.64	0.64	**0.38**	**0.33**	**0.31**	**0.29**	**0.26**	**0.28**	**0.25**	–	**0.23**	**0.27**	**0.39**	0.32	0.29	0.31	0.32	0.33	0.33	0.36	0.41	0.43	0.41	0.54	0.55	0.64	0.65	0.68	0.72	0.76	0.74
-10	0.81	0.83	0.80	0.79	0.77	0.79	0.68	0.69	**0.39**	**0.36**	**0.35**	**0.30**	**0.30**	**0.30**	**0.28**	**0.23**	–	**0.26**	**0.37**	0.32	0.29	0.31	0.32	0.31	0.34	0.36	0.41	0.44	0.42	0.54	0.53	0.65	0.65	0.67	0.71	0.75	0.75
-5	0.83	0.83	0.83	0.85	0.80	0.77	0.52	0.54	0.42	**0.39**	**0.37**	**0.36**	**0.34**	**0.31**	**0.26**	**0.27**	**0.26**	–	**0.36**	0.25	0.25	0.26	0.32	0.35	0.36	0.40	0.39	0.44	0.44	0.58	0.56	0.70	0.70	0.71	0.75	0.79	0.79
0	0.96	1.00	0.94	0.96	0.91	0.91	0.62	0.66	0.58	0.58	0.53	0.45	0.44	0.50	**0.44**	**0.39**	**0.37**	**0.36**	–	**0.35**	**0.38**	**0.38**	**0.40**	**0.43**	0.47	0.48	0.58	0.63	0.60	0.76	0.67	0.78	0.77	0.81	0.83	0.83	0.84
5	0.82	0.81	0.83	0.84	0.81	0.78	0.51	0.56	0.48	0.45	0.43	0.39	0.38	0.35	0.32	0.32	0.32	0.25	**0.35**	–	**0.28**	**0.29**	0.32	0.32	0.35	0.40	0.38	0.45	0.43	0.57	0.56	0.69	0.67	0.69	0.73	0.77	0.79
10	0.85	0.86	0.81	0.85	0.81	0.77	0.50	0.55	0.43	0.41	0.38	0.37	0.35	0.33	0.30	0.29	0.29	0.25	**0.38**	**0.28**	–	**0.24**	**0.30**	**0.27**	**0.31**	**0.33**	**0.37**	**0.40**	**0.39**	0.53	0.49	0.75	0.80	0.79	0.80	0.80	0.78
15	0.80	0.81	0.84	0.82	0.81	0.79	0.65	0.68	0.43	0.39	0.35	0.38	0.35	0.31	0.29	0.31	0.31	0.26	**0.38**	**0.29**	**0.24**	–	**0.28**	**0.27**	**0.30**	**0.31**	**0.35**	**0.41**	**0.37**	**0.52**	0.54	0.70	0.69	0.69	0.76	0.79	0.78
20	0.80	0.82	0.82	0.80	0.79	0.78	0.67	0.68	0.40	0.40	0.37	0.38	0.37	0.35	0.31	0.32	0.32	0.32	**0.40**	**0.32**	**0.30**	**0.28**	–	**0.26**	**0.25**	**0.30**	**0.34**	**0.39**	**0.37**	**0.47**	0.52	0.64	0.63	0.66	0.70	0.75	0.75
25	0.77	0.81	0.75	0.70	0.68	0.67	0.55	0.58	0.43	0.41	0.41	0.37	0.36	0.35	0.35	0.33	0.31	0.35	**0.43**	**0.32**	**0.27**	**0.27**	**0.26**	–	**0.27**	**0.30**	**0.28**	**0.34**	**0.32**	**0.47**	**0.49**	0.75	0.79	0.79	0.79	0.80	0.79
30	0.89	0.93	0.89	0.87	0.85	0.85	0.70	0.70	0.41	0.43	0.40	0.41	0.36	0.37	0.35	0.33	0.34	0.36	0.47	0.35	**0.31**	**0.30**	**0.25**	**0.27**	–	**0.27**	**0.30**	**0.33**	**0.32**	**0.46**	**0.45**	0.71	0.76	0.79	0.80	0.82	0.82
35	0.79	0.79	0.72	0.67	0.66	0.68	0.57	0.61	0.43	0.42	0.41	0.41	0.39	0.39	0.39	0.36	0.36	0.40	0.48	0.40	**0.33**	**0.31**	**0.30**	**0.30**	**0.27**	–	**0.29**	**0.31**	**0.29**	**0.43**	**0.46**	0.72	0.75	0.78	0.77	0.79	0.81
40	0.87	0.92	0.81	0.87	0.81	0.80	0.55	0.61	0.46	0.43	0.43	0.44	0.42	0.40	0.40	0.41	0.41	0.39	0.58	0.38	**0.37**	**0.35**	**0.34**	**0.28**	**0.30**	**0.29**	–	**0.30**	**0.27**	**0.44**	**0.45**	0.72	0.78	0.81	0.78	0.81	0.79
45	0.81	0.85	0.77	0.73	0.68	0.69	0.57	0.64	0.47	0.47	0.48	0.47	0.44	0.46	0.44	0.43	0.44	0.44	0.63	0.45	**0.40**	**0.41**	**0.39**	**0.34**	**0.33**	**0.31**	**0.30**	–	**0.29**	**0.45**	**0.49**	0.78	0.78	0.80	0.79	0.81	0.82
50	0.77	0.83	0.72	0.69	0.68	0.68	0.55	0.63	0.47	0.46	0.46	0.43	0.43	0.41	0.40	0.41	0.42	0.44	0.60	0.43	**0.39**	**0.37**	**0.37**	**0.32**	**0.32**	**0.29**	**0.27**	**0.29**	–	**0.38**	**0.46**	0.74	0.75	0.77	0.76	0.79	0.79
55	0.82	0.88	0.85	0.86	0.84	0.86	0.69	0.76	0.61	0.57	0.59	0.58	0.57	0.57	0.53	0.54	0.54	0.58	0.76	0.57	0.53	**0.52**	**0.47**	**0.47**	**0.46**	**0.43**	**0.44**	**0.45**	**0.38**	–	**0.43**	0.71	0.78	0.81	0.80	0.79	0.76
60	0.81	0.81	0.78	0.78	0.72	0.77	0.68	0.71	0.57	0.57	0.58	0.55	0.56	0.57	0.53	0.55	0.53	0.56	0.67	0.56	0.49	0.54	0.52	**0.49**	**0.45**	**0.46**	**0.45**	**0.49**	**0.46**	**0.43**	–	**0.53**	**0.52**	**0.55**	**0.61**	0.69	**0.70**
65	0.70	0.74	0.73	0.72	0.67	0.67	0.67	0.72	0.66	0.71	0.67	0.66	0.66	0.72	0.65	0.64	0.65	0.70	0.78	0.69	0.75	0.70	0.64	0.75	0.71	0.72	0.72	0.78	0.74	0.71	**0.53**	–	**0.35**	**0.41**	**0.50**	**0.65**	**0.63**
70	0.68	0.72	0.72	0.67	0.67	0.67	0.64	0.70	0.66	0.68	0.69	0.66	0.66	0.67	0.66	0.65	0.65	0.70	0.77	0.67	0.80	0.69	0.63	0.79	0.76	0.75	0.78	0.78	0.75	0.78	**0.52**	**0.35**	–	**0.29**	**0.41**	**0.54**	**0.54**
75	0.72	0.71	0.71	0.68	0.69	0.72	0.70	0.70	0.68	0.73	0.67	0.67	0.69	0.71	0.69	0.68	0.67	0.71	0.81	0.69	0.79	0.69	0.66	0.79	0.79	0.78	0.81	0.80	0.77	0.81	**0.55**	**0.41**	**0.29**	–	**0.35**	**0.52**	**0.54**
80	0.72	0.75	0.73	0.67	0.70	0.73	0.69	0.71	0.73	0.77	0.75	0.72	0.75	0.74	0.73	0.72	0.71	0.75	0.83	0.73	0.80	0.76	0.70	0.79	0.80	0.77	0.78	0.79	0.76	0.80	**0.61**	**0.50**	**0.41**	**0.35**	–	**0.48**	**0.49**
85	0.70	0.72	0.70	0.68	0.68	0.75	0.65	0.68	0.79	0.83	0.80	0.79	0.79	0.80	0.78	0.76	0.75	0.79	0.83	0.77	0.80	0.79	0.75	0.80	0.82	0.79	0.81	0.81	0.79	0.79	0.69	**0.65**	**0.54**	**0.52**	**0.48**	–	**0.36**
90	0.67	0.72	0.74	0.71	0.71	0.76	0.71	0.70	0.80	0.84	0.78	0.80	0.79	0.80	0.77	0.74	0.75	0.79	0.84	0.79	0.78	0.78	0.75	0.79	0.82	0.81	0.79	0.82	0.79	0.76	0.70	**0.63**	**0.54**	**0.54**	**0.49**	**0.36**	–

Table 1. The shape metric between pairs of arbitrary views. Costs in bold in each row show the region where the metric satisfies criteria 1 and 2, *i.e.*, is monotonic and globally selective of the prototype.

of the neighborhood δ over which this holds is greater for d_s than for d_c, Figure 4. More significantly, however, the shock graph matching metric brings out important differences arising from a slight change in viewpoint. Figure 4 illustrates this by plotting the distance between two adjacent views, for each view, as a function of viewing angle, *i.e.*, an estimate of the derivative of the metric with viewpoint changes at each point. Observe how the shock metric d_s depicts significant regions of similarity, noted by 1 in Figure 4, while maintaining noticeable differences which give rise to aspect boundaries, noted by 2 in Figure 4. The curve matching metric, on the other hand, misses some of the sudden changes in shape with slight changes in view.

Thus, while it is possible to use d_c with the current approach, despite the additional computation cost, we selected d_s, mainly for its larger extent of monotonicity with viewpoint variation, and its superior sensitivity to significant view changes between adjacent views.

2.2 Generating Aspects and Prototypes

The generation of aspects can be viewed as a segmentation problem, not unlike that of intensity images, but now applied to the viewing sphere endowed with a similarity metric. Two approaches can be adopted in analogy to approaches of segmentation of intensity images: (i) group similar views, a region-growing approach, and (ii) detect significant peaks in local differences, an edge-based ap-

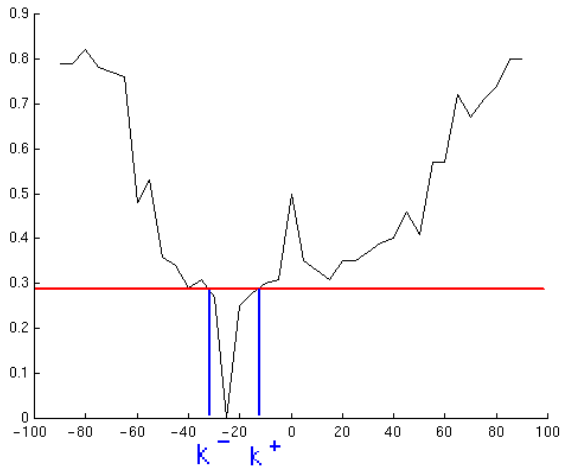

Figure 5. Criterion 2 constrains the upper limits of the boundary views k^+ and k^- of an aspect for each view beyond the monotonic region. k^+ and k^- are the views most distant from the prototype candidate which remain in the monotonic region and which are below the value of the global minimum outside this region.

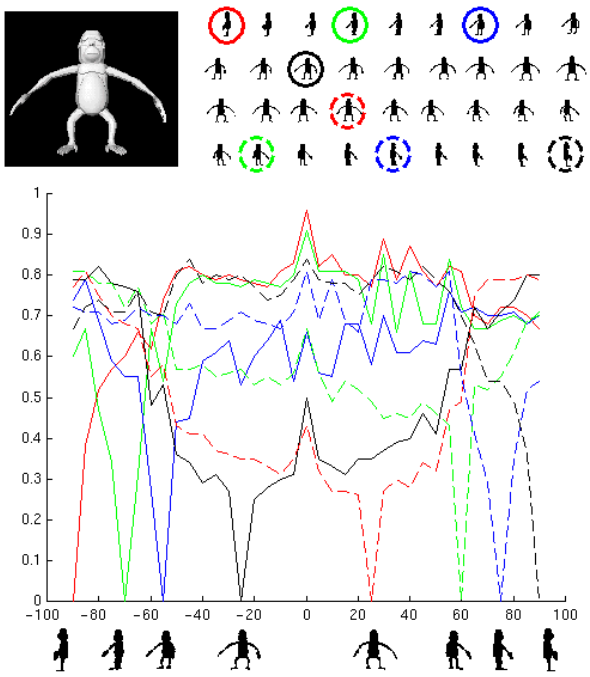

Figure 6. (Top) The original 3D model "Ape", left, generates the set of views on the right. Each circled view is a sample view for which the monotonicity condition is examined. We selected prototypical views for our sample, but the same condition holds for all other views. (Bottom) The set of 8 graphs for each choice of views m above $d(V_n^m, V_n^{m+i})$ is plotted as a function of i. Observe the local neighborhood where d is monotonically increasing away from the base view.

proach. While the distinct peaks in Figure 4 can potentially provide for an edge-based approach to generating aspects, that changes can sometimes be diffused, prompts us to adopt a region-growing approach. We must first place additional constraints on defining an aspect, however. While the monotonicity condition limits the size of an aspect centered around a candidate prototype, the metric of similarity should also be able to differentiate between aspect and non-aspect views:

Criteria 2 (Object-Specific Distinctiveness of Aspect Views)*: For each aspect A_n^m with prototypical view V_n^m, we must have $d(V_n^i, V_n^m) < d(V_n^j, V_n^m)$ for any $V_n^i \in A_n^m$, and $V_n^j \notin A_n^m$*

Observe that this delimits the possible aspect boundaries for each candidate prototype views, since for each V_n^m we

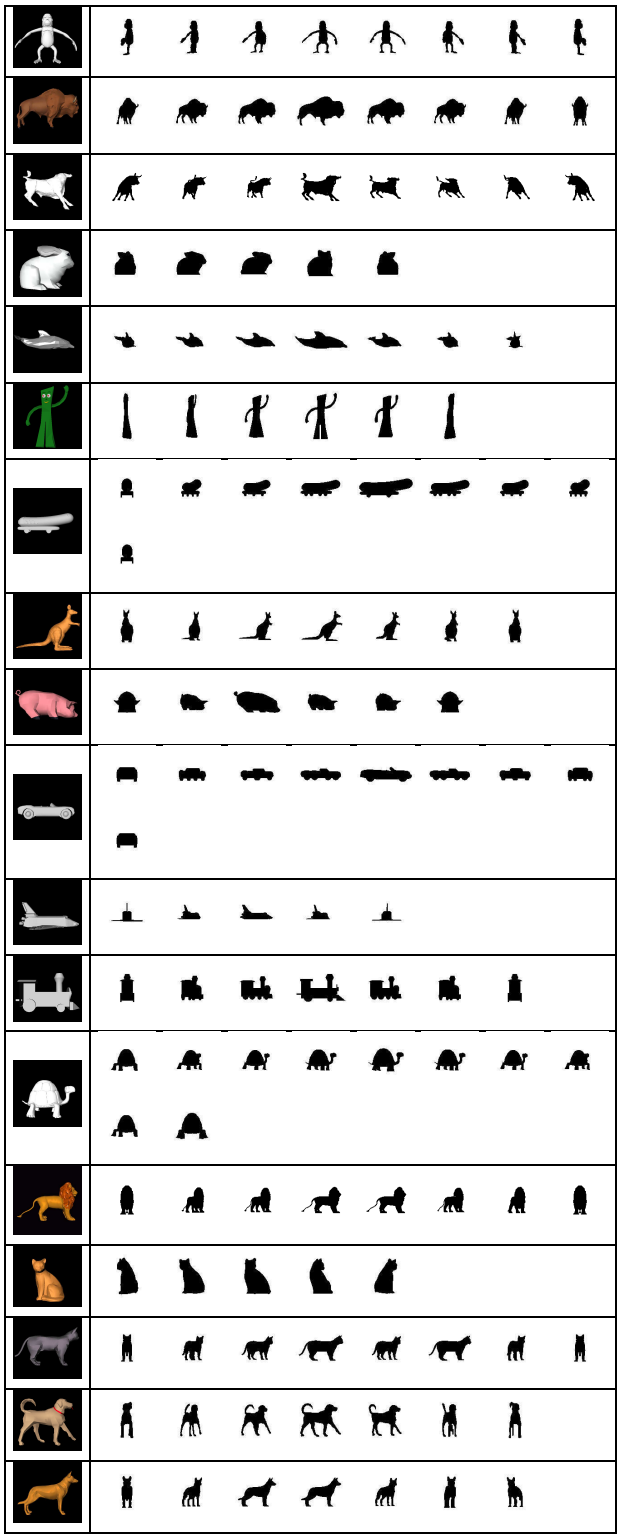

Figure 7. The set of prototypical views of various models included in our database

must have

$$\max_{V_n^i \in A_n^m} d(V_n^i, V_n^m) < \min_{V_n^j \notin A_n^m} d(V_n^j, V_n^m), \qquad (2)$$

or, by local monotonicity condition,

$$\max\left(d(V_n^{k^+}, V_n^m), d(V_n^{k^-}, V_n^m)\right) < \min_{V_n^j \in A_n^m} d(V_n^j, V_n^m) \qquad (3)$$

This in turn defines upper bounds $V_n^{k^+}$ and $V_n^{k^-}$ for each view V_n^m, namely, the first view in either direction of view angle variations where $d(V_n^{k^+}, V_n^m)$ reaches the value of the global minimum outside the monotonic region around each view, Figure 5. Clearly, views closer to the prototype then k^+ and k^- also satisfy this condition, but we are seeking the largest aspect which implies extending the aspect in the monotonic region until the global minimum beyond the monotonic region is reached. This places limits on the boundary of each aspect for each view.

Table 1 illustrates the matrix of metric differences between all the views of an object. Each row (or column) gives rise to a graph like those depicted in Figure 5; the largest possible aspect for each view is depicted in bold. Note that only the numbers in bold are of interest, thus there is room for future improvements in efficiency to take advantage of this fact.

The aspects and prototypes are then generated by seeking a set maximal region where Criteria 1 and 2 hold. We employ a "seeded region-growing" algorithm [1] to maximize the size of each aspect competitively, starting from each view, which is considered a seed. The two views with the minimal distance are merged into an aspect and represented by a prototype if they do not exceed the upper bounds. The procedure is repeated by replacing the merged views with a representative prototype, which is reselected in each iteration to minimize the maximum distance within the aspect. The algorithm converges with no parameters and leads to intuitive prototypes and aspects.

3 3D Recognition by Matching 2D Views

The previous section described how aspects and prototypes for each object can be generated such that *(i)* the distance between each view in an aspect and that aspect's prototype is less than this view's distance to any other aspect prototype, *(ii)* the number of aspects was reduced by grouping to the largest clusters not violating criteria 1 and 2, namely monotonicity and object-specific view distinctiveness. We now examine indexing into a set of 3D objects represented by their aspects and prototypical views.

Consider an unknown view of unknown object V_j^i. We match this view to all prototypical views V_n^m of all objects

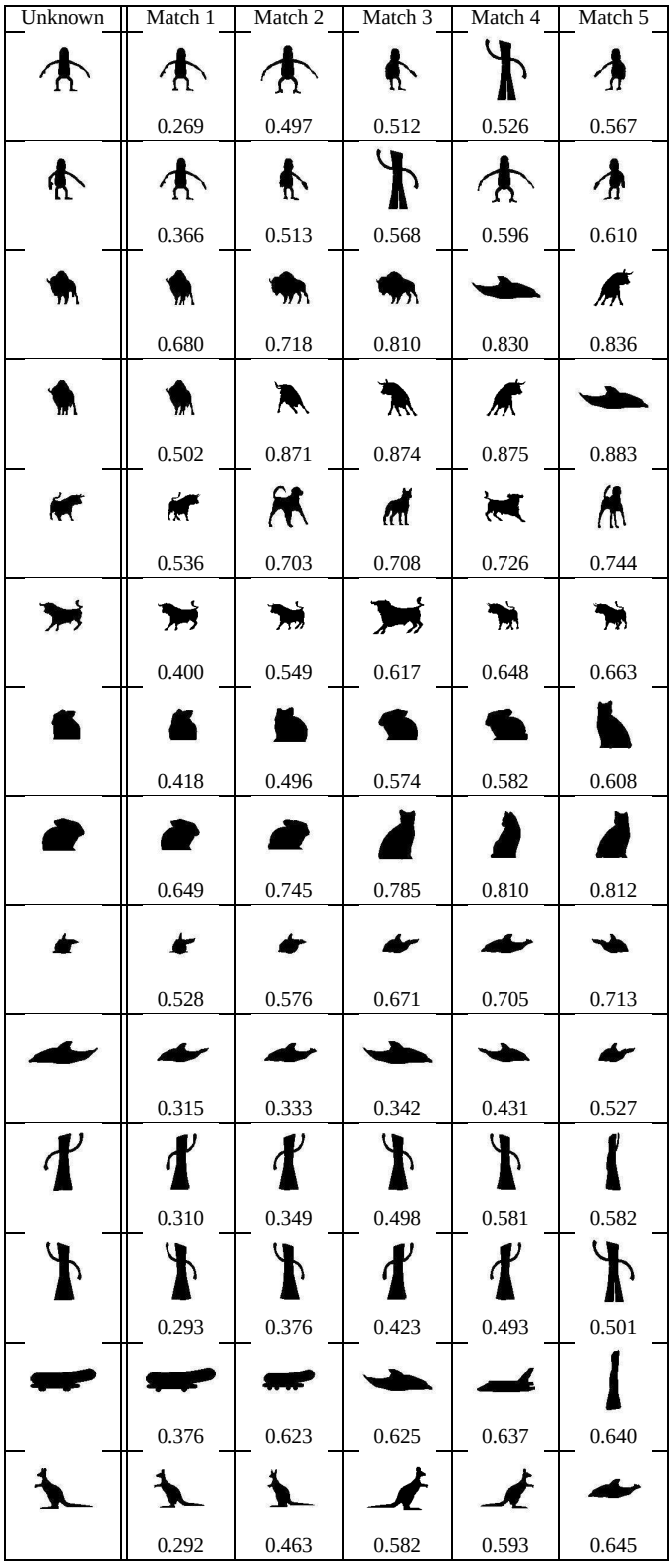

Figure 8. The result of matching. Unknown views of unknown objects are matched against stored prototypes and ordered by increasing distance.

in the database, and order the set of matched prototypes in order of increasing metric. The top choice gives the identity of the 3D object and its pose. The following theorem shows that if the objects in the database are distinct enough, the similarity-based aspect matching method proposed here will find the correct object.

Definition 1 (Aspect Diameter)*: The diameter of an aspect A_n^m is twice the maximum distance of the view in the aspect from the prototype, i.e.,*

$$D_n^m(V_n^m) = 2 \max_{V_n^i \in A_n^m} d(V_n^i, V_n^m) = 2 \max_{\pm} d(V_n^{k^\pm}, V_n^m).$$

Definition 2 (Aspect-Separable)*: Consider a database of N objects $\{O_1, ..., O_N\}$ with views $\{V_n^1, ...V_n^M\}$ with associated prototype $\{V_n^m\}$ satisfying criteria 1 and 2. The database is* aspect-separable if for any pair of distinct prototypes V_n^m and V_p^q we have

$$d(V_n^m, V_p^q) > D(V_n^m) \tag{4}$$

Theorem 1 *Given an unknown view V_j^i, selected from an aspect-separable database, the object identified by the shape metric is the correct one, i.e., $j = \arg\min\ d(V_j^i, V_n^m)$.*

Proof 1 Without loss of generality, let $V_n^{k^+}$ be the most distant view from V_n^m in aspect A_n^m such that $D_n^m = 2d(V_n^m, V_n^{k^+})$. By the triangle inequality and monotonicity

$$\begin{aligned} d(V_n^m, V_p^q) &\leqslant d(V_n^m, V_n^i) + d(V_n^i, V_p^q) \\ &\leqslant d(V_n^m, V_n^{k^+}) + d(V_n^i, V_p^q). \end{aligned} \tag{5}$$

Since, by hypothesis

$$2d(V_n^m, V_n^{k^+}) = D_n^m < d(V_n^m, V_p^q), \tag{6}$$

we have by combining equation (5) and (6) and subtracting $d(V_n^m, V_p^q)$

$$d(V_n^m, V_n^{k^+}) < d(V_n^i, V_p^q).$$

Finally, since $d(V_n^i, V_n^m) \leqslant d(V_n^{k^+}, V_n^m)$,
we have

$$d(V_n^i, V_n^m) < d(V_n^i, V_p^q).$$

for all prototypes V_p^q distinct from V_n^m. ■

Intuitively, if the objects in the database produce views which are distinct enough, as measured by the distance between the two most distant views in an aspect, the objects can be distinguished by comparing the similarity between their prototypical views. On the other hand, the introduction of an object with views rather similar to those of another object in the database creates an ambiguity. In this case, the above analysis suggests grouping of similar objects into categories.

It should be noted that Theorem 1 establishes a sufficient condition, but not a necessary one. Thus, in the construction of a 3D object database, aspects and their prototypes are first generated for each novel object. These prototypes are compared against existing prototypical views to verify aspect-separability. If the Inequality (1) is violated, it is still possible for the similarity based matching to produce the correct object, but this is no longer guaranteed via Theorem 1. On the other hand, if the Inequality (1) is satisfied, then we are assured that recognition by similarity-based matching produces the correct result.

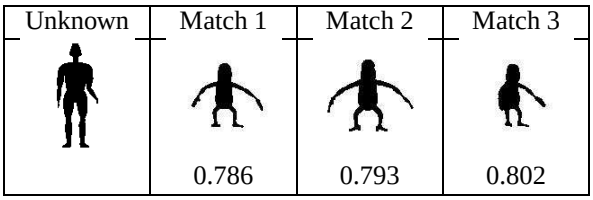

Figure 9. Results of matching a view of an object not in our database with the prototypical views of objects in the database.

4 Results

We generated views at random angles of randomly selected objects from our database of models. Each unknown view was matched against the prototypical views in the database. The results from matching all the prototypical views and the unknown image are sorted, and the best match gives the unknown object. Figure 8 shows the top 5 matches and their match costs for several examples. Table 2 summarizes the results. The best match is always correct here since the objects in the database are sufficiently distinct.

We also tested the performance of the recognition scheme for views of models which were not in our database, but which were similar in some way to an object in the database. For example, the "robot" model was somewhat close to the "ape" figure; Figure 9 shows that the top match gives the correct object and pose.

While these results are promising, we recognize the need to test the approach on a much larger database. In fact, since the paper was submitted we have extended the database to include 65 objects with similar results. The main limiting factor is the time required to compute a shock match. Since each match currently takes around 5 minutes on an SGI Indigo II 195Mhz, performing a large number of comparisons

	Best Match 1	Best Match 2	Best Match 3	Best Match 4	Best Match 5
Object Correctly Detected	100.0%	87.7%	66.6%	60.0%	53.3%

Table 2. Results of matching Unknowns

can be time consuming. We are considering a number of items which would improve time efficiency. In addition, the data files resulting from the matches take on average 1 MB of disk space, which can become a restriction when matching large datasets. These are limitations which can be drastically reduced by improving the shock matching code. Also, while "ground views" are the more interesting of the views, the full set of views should be incorporated by applying the region-growing algorithm to the viewing sphere endowed with the shape metric. The results presented here encourage further development of this approach.

Acknowledgements: Benjamin Kimia acknowledges the support of NSF grants IRI-9700497 and BCS-9980091.

References

[1] R. Adams and L. Bischof. Seeded region growing. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16(6):641–647, 1994.

[2] A. Barr. Superquadrics and angle-preserving transformations. *IEEE CGA*, 1(1):11–23, January 1981.

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

[4] K. Bowyer and C. Dyer. Aspect graphs: an introduction and survey of recent results. *IJIST*, 2:315–328, 1990.

[5] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models: Their training and application. *CVIU*, 61(1):38–59, January 1995.

[6] F. Cutzu and M. Tarr. Inferring perceptual saliency fields from viewpoint-dependent recognition data. *NeurComp*, 11:1331–1348, 1999.

[7] C. M. Cyr, A. F. Kamal, T. B. Sebastian, and B. B. Kimia. 2D-3D registration based on shape matching. In *Proceedings of Mathematical Methods in Biomedical Image Analysis*, pages 198–203, 2000.

[8] S. Dickinson, A. Pentland, and A. Rosenfeld. 3D shape recovery using distributed aspect matching. *PAMI*, 14(2):174–198, February 1992.

[9] D. Eggert and K. Bowyer. Computing the perspective projection aspect graph of solids of revolution. *PAMI*, 15(2):109–128, February 1993.

[10] D. Eggert, K. Bowyer, C. Dyer, H. Christensen, and D. Goldgof. The scale space aspect graph. *PAMI*, 15(11):1114–1130, November 1993.

[11] K. Ikeuchi and T. Kanade. Automatic generation of object recognition programs. *PIEEE*, 76(8):1016–1035, August 1988.

[12] P. Klein, T. Sebastian, and B. Kimia. Shape matching using edit-distance: an implementation. In *Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 781–790, Washington, D.C., January 7-9 2001.

[13] J. J. Koenderink and A. J. van Doorn. The singularilarities of the visual mapping. *Biol. Cyber.*, 24:51–59, 1976.

[14] J. J. Koenderink and A. J. van Doorn. The internal representation of solid shape with respect to vision. *Biol. Cyber.*, 32:211–216, 1979.

[15] D. Kriegman and J. Ponce. Computing exact aspect graphs of curved objects: Solids of revolution. *IJCV*, 5(2):119–135, November 1990.

[16] S. Nayar, S. Rene, and H. Murase. Realtime 100 object recognition system. In *Proceedings 1996 IEEE International Conference on Robotics and Automation*, pages 2321–2325, 1996.

[17] A. Pentland and S. Sclaroff. Closed-form solutions for physically based shape modeling and recognition. *PAMI*, 13(7):715–729, July 1991.

[18] T. B. Sebastian, J. J. Crisco, P. N. Klein, and B. B. Kimia. Constructing 2D curve atlases. In *Proceedings of Mathematical Methods in Biomedical Image Analysis*, pages 70–77, 2000.

[19] T. B. Sebastian, P. N. Klein, and B. B. Kimia. Alignment-based recognition of shape outlines. In *IWVF*, pages Accepted –To Appear, 2001.

[20] T. B. Sebastian, P. N. Klein, and B. B. Kimia. Recognition of shapes by editing shock graphs. In *ICCV*, pages Accepted – To appear, 2001.

[21] I. Shimshoni and J. Ponce. Finite-resolution aspect graphs of polyhedral objects. *PAMI*, 19(4):315–327, April 1997.

[22] J. Stewman and K. Bowyer. Creating the perspective projection aspect graph of polyhedral objects. volume 9, pages 494–500, 1988.

[23] J. Subrahmonia, D. B. Cooper, and D. Keren. Practical Reliable Bayesian Recognition of 2D and 3D Objects Using Implicit Polynomials and Algebraic Invariants. Technical Report LEMS-107, Brown University, May 1992.

[24] S. Ullman and R. Basri. Recognition by linear combinations of models. *IEEE*, 13(10):992–1006, 1991.

[25] D. Weinshall and M. Werman. On view likelihood and stability. *PAMI*, 19(2):97–108, February 1997.