View-Based 3-D Object Recognition using Shock Graphs*

Diego Macrini
Department of Computer Science
University of Toronto

Ali Shokoufandeh
Department of Mathematics and Computer Science
Drexel University

Sven Dickinson
Department of Computer Science
University of Toronto

Kaleem Siddiqi
Centre for Intelligent Machines
School of Computer Science
McGill University

Steven Zucker
Center for Computational Vision and Control
Yale University

Abstract

The shock graph is an emerging shape representation for object recognition, in which a 2-D silhouette is decomposed into a set of qualitative parts, captured in a directed acyclic graph. Although a number of approaches have been proposed for shock graph matching, these approaches do not address the equally important indexing problem. We extend our previous work in both shock graph matching and hierarchical structure indexing to propose the first unified framework for view-based 3-D object recognition using shock graphs. The heart of the framework is an improved spectral characterization of shock graph structure that not only drives a powerful indexing mechanism (to retrieve similar candidates from a large database), but also drives a matching algorithm that can accommodate noise and occlusion. We describe the components of our system and evaluate its performance using both unoccluded and occluded queries. The large set of recognition trials (over 25,000) from a large database (over 1400 views) represents one of the most ambitious shock graph-based recognition experiments conducted to date.

1 Introduction

There are two approaches to 3-D object recognition. One assumes a 3-D object-centered model, and attempts to match 2-D image features to viewpoint-invariant 3-D model features. Over the last decade, this approach has given way to a viewer-centered approach, where the 3-D model is replaced by a collection of 2-D views. These views can be represented in terms of segmented features, such as lines or regions, or in terms of the photometric "appearance" of the object. Although these latter, appearance-based recognition schemes have met with great success, it must be understood that they address the task of exemplar-based recognition. When faced with novel exemplars belonging to known classes, they simply do not scale up.

To achieve such categorical, or generic, object recognition requires a representation that is invariant to within-class shape deformations. One such powerful representation is offered by the shock graph [11], which represents the silhouette of an object in terms of a set of qualitatively defined parts, organized in a hierarchical, directed acyclic graph. Figure 1 illustrates an example of a two-dimensional shape, its shocks (singularities), and the resulting shock graph. In previous work, we introduced the first algorithm for matching two shock graphs, and showed that it could be used to recognize novel exemplars from known classes [10]. Since then, other approaches to shock graph matching have emerged, including [6] and [7]. However, earlier approaches, including our own, have not been extensively tested on noisy graphs, occluded scenes, or cluttered scenes.

A shock graph representation of shape suggests the use of graph matching techniques for shape recognition. However, matching (graph or otherwise) is only half the problem. Without an effective *indexing* mechanism with which to narrow a large database down to a small number of candidates, recognition degenerates to matching a query to each model in the database. In the case of view-based object recognition, in which a large number of objects map to an

*The authors would like to acknowledge the programming support of Maxim Trokhimtchouk, Carlos Phillips, and Pavel Dimitrov. The authors would also like to express their thanks to Norio Katayama for the use of their SR-tree implementation. Finally, the authors would like to acknowledge the generous support of NSERC, FCAR, CFI, CITO, and NSF.

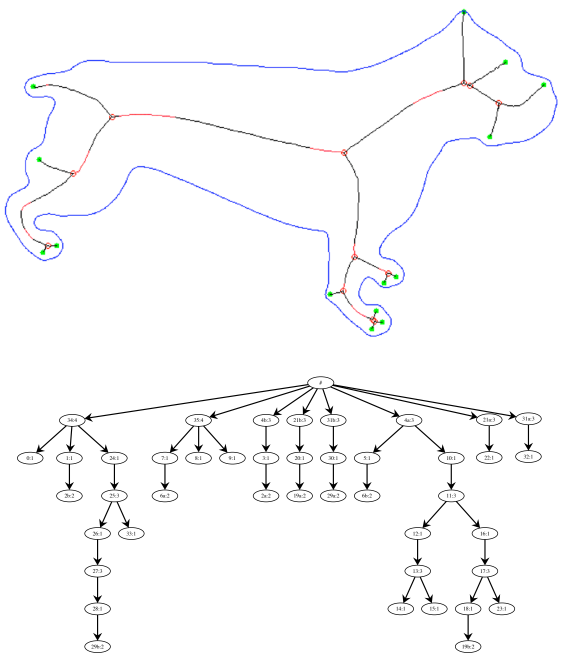


Figure 1. A Two-Dimensional Shape and its Corresponding Shock Graph. The nodes represent groups of singularities (shocks) along with their geometric attributes. The edges are between adjacent shock groups and in a direction opposite to Blum's grassfire flow [10].

even larger number of views, such a linear search is intractable. Unfortunately, very few researchers, in either the computer vision or graph algorithms communities, have addressed the important problem of graph indexing. How, then, can we exploit the power of the shock graph to perform view-based object recognition?

In recent work, we introduced a novel indexing method which maps the structure of a directed acyclic graph to a point in low-dimensional space [9]. This same mapping, in fact, was used as the basis for our shock graph matching algorithm [10]. Using standard, nearest-neighbor search methods, this compact, structural signature was used to retrieve structurally similar candidates from a database. The highest scoring candidates, in turn, were compared to the query using our matching algorithm, with the "closest" candidate used to "recognize" the object. Our experiments showed that the target ranked highly among the candidates, even in the presence of noise and occlusion.

Armed with a unified approach to the indexing and matching of graphs, we now turn to the problem of view-based object recognition using shock graphs. In fact, we are not the first to apply shock graphs to this problem. In recent work, Cyr and Kimia [2, 7] explore the important problem of how to partition the view sphere of a 3-D object using a collection of shock graphs. However, they do not address the shock graph indexing problem, resorting to a linear search of all views in the database in order to recognize an object. Even for small object databases, the number of views required per object renders this approach intractable. In this paper, we unify our shock graph indexing and matching techniques to yield a novel, effective method for view-based 3-D object recognition.

2 A Compact Encoding of Graph Structure

In [10], we introduced a transformation mapping the structure of a directed acyclic graph to a point in low-dimensional space. As mentioned earlier, this mapping was the heart of an algorithm for matching two *shock trees*, derivable from shock graphs in linear time. This same transformation later gave rise to an indexing mechanism, which used the low-dimensional, structural signature of a shock tree to select structurally similar candidates from a database of shock trees [9]. In this latter paper, we analyzed the stability of a tree's signature to certain restricted classes of perturbations.

In a recent paper on matching multi-scale image decompositions, we have strengthened this encoding from undirected, unique rooted trees to directed acyclic graphs, yielding a more powerful characterization of graph structure [8]. This new formulation has led to a broader stability analysis that accommodates *any* graph perturbation in terms of node addition and/or deletion. Furthermore, we extend our matching algorithm to deal with directed acyclic graphs rather than undirected, unique rooted trees.

To encode the structure of a DAG, we turn to the domain of eigenspaces of graphs, first noting that any graph can be represented as an antisymmetric $\{0, 1, -1\}$ adjacency matrix, with 1's (-1's) indicating a forward (backward) edge between adjacent nodes in the graph (and 0's on the diagonal). The eigenvalues of a graph's adjacency matrix encode important structural properties of the graph, and are stable under minor perturbations in structure. Our goal, therefore, is to map the eigenvalues of a DAG to a point in some low-dimensional space, providing a stable, compact encoding of structure.

Specifically, let T be a DAG whose maximum branching factor is $\Delta(T)$, and let the subgraphs of its root be $T_1, T_2, \ldots, T_{\delta(T)}$. For each subgraph, T_i, whose root degree is $\delta(T_i)$, we compute[1] the magnitudes of the eigenvalues of T_i's submatrix, sort them in decreasing order by absolute value, and let S_i be the sum of the $\delta(T_i) - 1$ largest absolute values. The sorted S_i's become the components

[1] We use SVD to compute the magnitudes of the eigenvalues.

of a $\Delta(T)$-dimensional vector assigned to the DAG's root. If the number of S_i's is less than $\Delta(T)$, then the vector is padded with zeroes. We can recursively repeat this procedure, assigning a vector to each nonterminal node in the DAG, computed over the subgraph rooted at that node. We call each such vector a *topological signature vector*, or TSV. The details of this transformation, the motivation for each step, and an evaluation of its properties is given in [8].

3 Shock Graph Indexing

Given a query shape, represented by a shock graph, the goal of indexing is to efficiently retrieve, from a large database, similar shock graphs that might account for the query or some portion thereof (in the case of an occluded query or a query representing a cluttered scene). These *candidate* model graphs will then be compared directly with the query, i.e., *verified*, to determine which candidate model best accounts for the query. We therefore seek an effective, low-dimensional index for shock graph recognition that captures both local and global structural properties, is invariant to re-ordering of a node's descendants, has low ambiguity, is stable to minor perturbations of graph structure, and is efficiently computed.

Our topological signature vector, in fact, satisfies these six criteria, suggesting that a model DAG's structure can be represented as a vector in δ-dimensional space, where δ is an upper bound on the degree of any vertex of any image or model DAG. If we could assume that an image DAG represents a properly segmented, unoccluded object, then the TSV computed at the query DAG's root, could be compared with those topological signature vectors representing the roots of the model DAGs. The vector distance between the image DAG's root TSV and a model DAG's root TSV would be inversely proportional to the structural similarity of their respective DAGs, as finding two subgraphs with "close" eigenvalue sums represents an approximation to finding the largest subgraph isomorphism.

Unfortunately, this simple framework cannot support either cluttered scenes or large occlusion, both of which result in the addition or deletion of significant structure. In either case, altering the structure of the DAG will affect the TSV's computed at its nodes. The signatures corresponding to the roots of those subgraphs (DAGs) that survive the occlusion will not change. We can accommodate such perturbations through a local indexing framework analogous to that used in a number of geometric hashing methods, e.g., [4]. Rather than storing a model DAG's root signature, we will store the signatures of *each* node in the model DAG. At each such point (node signature) in the database, we will associate a pointer to the object model containing that node as well as a pointer to the corresponding node in the model DAG (allowing access to node label information). Since a given model subgraph can be shared by other model DAGs, a given signature (or location in δ-dimensional space) will point to a list of (model object, model node) ordered pairs.

Each node in the query DAG will generate a set of (model object, model node) votes. To collect these votes, we set up an accumulator with one bin per model object. Furthermore, we can weight the votes that we add to the accumulator according to two important factors. Given a query node and a model node (retrieved from the database), we weight the vote according to the distance between their respective TSV's – the closer the signatures, the more weight the vote gets. Furthermore, we weight the vote according to the complexity of its corresponding subgraph, allowing larger and more complex subgraphs (or "parts") to have higher weight. This can be easily accommodated within our eigenvalue framework, for the richer the structure, the larger its maximum eigenvalue [5]. Assembling the terms of our weight function yields:

$$W(q,m) = \frac{(1-\omega)||q||}{\mathcal{T}_q(1+||m-q||)} + \frac{\omega||m||}{\mathcal{T}_m(1+||m-q||)} \quad (1)$$

where q is the TSV of the query DAG node, m the TSV of the model DAG node (that is sufficiently close), $\mathcal{T}_q$ and $\mathcal{T}_m$ are the sums of the TSV norms of the entire query and model DAGs, respectively. The first term favors models that cover a larger proportion of the image, while the second favors models with more nodes accounted for, with convexity parameter ω.

Once the evidence accumulation is complete, those models whose support is sufficiently high are selected as candidates for verification. The bins can, in effect, be organized in a heap, requiring a maximum of $\Theta(\log k)$ operations to maintain the heap when evidence is added, where k is the number of non-zero object accumulators. Once the top-scoring models have been selected, they must be individually verified according to the matching algorithm described in the next section.

4 Shock Graph Matching

Our spectral characterization of graph structure forms the backbone of our indexing mechanism, as described in the previous section. Moreover, this *same* spectral characterization forms the backbone of our matching algorithm, thereby unifying the mechanisms of indexing and matching. In previous work [10], we showed that a shock graph could be transformed into a unique rooted *shock tree* in linear time. We introduced an algorithm for computing the distance between two shock trees (including correspondence) in the presence of noise and occlusion. We have recently strengthened our indexing and matching framework to include directed acyclic graphs [8], and have applied it to hierarchical blob matching. Space prohibits further discus-

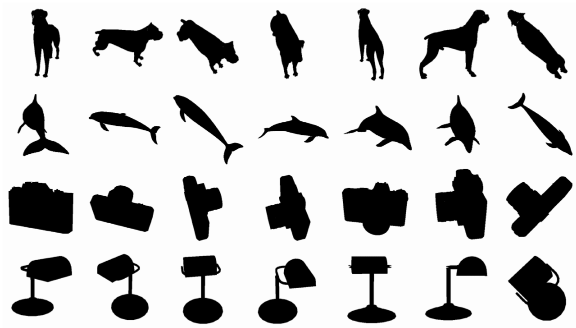


Figure 2. Some example object views drawn from our database.

sion of the algorithm, which can compute (in complexity better than $O(n^3)$) the distance between two DAGs, along with node correspondence, in the presence of noise and occlusion.

5 Experiments

We have systematically tested our integrated framework using both occluded and unoccluded queries. With over 27,000 trials and a database of over 1400 graphs, this represents one of the most comprehensive set of shock graph experiments to date. Our database consists of views computed from 3-D graphics models obtained from the public domain. Using a graphics modeling tool (3D Studio Max), each model is centered in a uniformly tessellated view sphere, and a silhouette is generated for each vertex in the tessellation. A shock graph is computed for each silhouette [3], and each node of the resulting graph is added to the model database, as described in Section 3. A sampling of the object views is shown in Figure 2.

In the first set of experiments, we evaluate the performance of the system on a set of unoccluded queries to an object view database. The database contains 1408 views describing 11 objects (128 uniformly sampled views per object). We then remove each view from the database and use it as a query to the remaining views. For each node of the query DAG, the indexing module will return all neighbors within a radius of 40% of the norm of (query) node. Evidence for models (containing a neighbor) is then accumulated, and the model bins are sorted. The indexer will return at most the highest scoring 50 candidates, but will return fewer if the sorted bins' contents drop suddenly. The candidates are matched to the query, using the matcher (see Section 4), and sorted according to similarity. If the query object (from which the query view was drawn) is the same as the model object from which the most similar candidate view is drawn, recognition is said to be successful, i.e., the object label is correct.[2]

Figure 3(a) plots recognition performance as a function of increasing number of objects (with 128 views per new object), while Figure 3(b) fixes the number of objects (11) and plots recognition performance as a function of sampling resolution. Recognition performance is very high, with better than 90% success until sampling resolution drops below 32 views (over the entire view sphere) per object. This demonstrates both the efficacy of the recognition framework and the viewpoint invariance of the shock graph, respectively. The most complex component of the algorithm is the matcher. However, with a fixed number (50) of verifications per query, independent of database size, complexity therefore varies as a function of nearest neighbor search and bin sorting, both of which are sublinear in the number of database views.

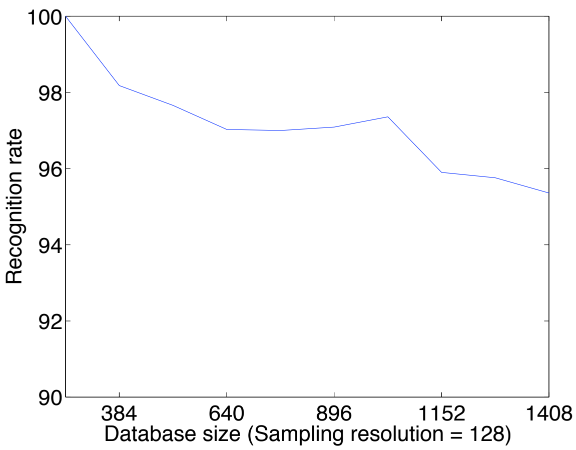
In the final experiment, shown in Figure 3(c), we plot recognition performance as a function of degree of occlusion (for the entire database) for occluded queries. To generate an occluded query, we randomly choose a node in the query DAG and delete the subgraph rooted at that node, provided that the node "mass" of the graph does not drop by more than 50%. As can be seen from the plot, performance decreases gradually as a function of occluder size (or, more accurately, the amount of "missing data"), reflecting the framework's ability to recognize partially visible objects.

It should be noted that in the above experiments, erroneous matches may be due to either ambiguous views (views shared by different objects) or to queries representing "degenerate" views, in which the removed view acting as a query was the last view of its class and therefore not expected to match other views on the object. Finally, space restrictions prohibit the presentation of our results on pose estimation.

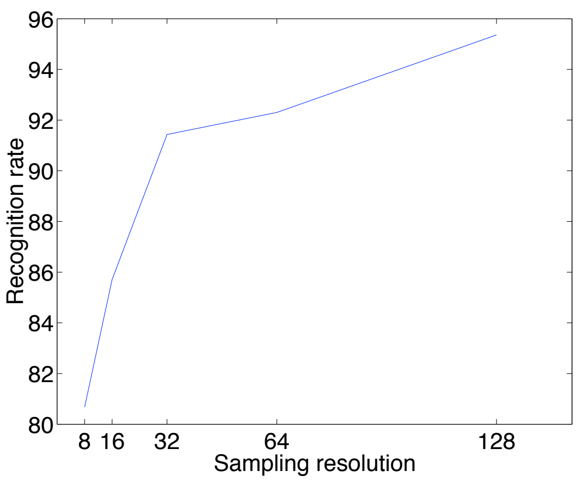
6 Conclusions

We have presented a unified mechanism for shock graph indexing and matching, and have applied it to the problem of view-based 3-D object recognition. Our spectral-based indexing framework quickly and effectively selects a small number of candidates, *including the correct one*, from a large database of model views from which our spectral-based matcher computes an accurate distance measure. Our scaling experiments demonstrate the framework's ability to effectively deal with large numbers of views, while our occlusion experiments establish its robustness. Current work is focused on view-cell clustering and strengthening the indexer to include more geometric and node label information. We also expect the results to improve when measures indicating the stability of nodes in the shock graphs are

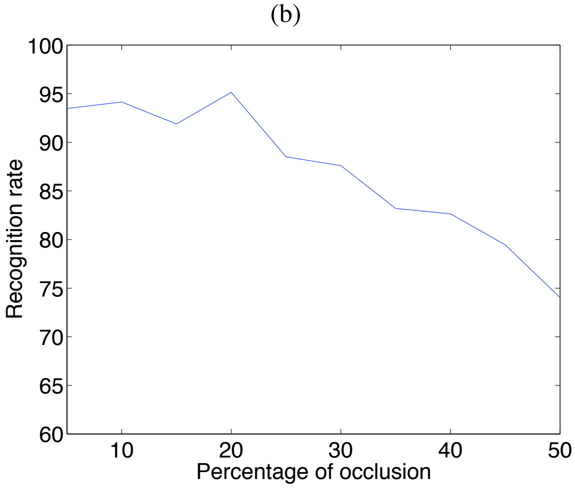
[2] Note that if multiple views (perhaps from different objects) are tied for "most similar", then each can be considered to be "most similar."



(a)



(b)



(c)

Figure 3. Recognition Performance: (a) Recognition performance as a function of object database size; (b) Recognition performance as a function of sampling resolution; and (c) Recognition performance as a function of degree of occlusion.

taken into account. In particular, it is known that nodes related to ligature are likely to be less stable and hence should be given less weight by both the indexer and the matcher [1].

References

[1] J. August, K. Siddiqi, and S. W. Zucker. Ligature instabilities in the perceptual organization of shape. *Computer Vision and Image Understanding*, 76(3):231–243, 1999.

[2] M. Cyr and B. Kimia. 3d object recognition using shape similarity-based aspect graph. In *Proceedings, ICCV*, Vancouver, B.C., 2001.

[3] P. Dimitrov, C. Phillips, and K. Siddiqi. Robust and efficient skeletal graphs. In *IEEE Conference on Computer Vision and Pattern Recognition*, Hilton Head, SC, June 2000.

[4] Y. Lamdan, J. Schwartz, and H. Wolfson. Affine invariant model-based object recognition. *IEEE Transactions on Robotics and Automation*, 6(5):578–589, October 1990.

[5] L. Lovász and J. Pelicán. On the eigenvalues of a tree. *Periodica Math. Hung.*, 3:1082–1096, 1970.

[6] M. Pelillo, K. Siddiqi, and S. Zucker. Matching hierarchical structures using association graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(11):1105–1120, November 1999.

[7] T. Sebastian, P. Klein, and B. Kimia. Recognition of shapes by editing shock graphs. In *Proceedings, ICCV*, Vancouver, B.C., 2001.

[8] A. Shokoufandeh, S. Dickinson, C. Jonsson, L. Bretzner, and T. Lindeberg. The representation and matching of qualitative shape at multiple scales. In *Proceedings, ECCV*, Copenhagen, May 2002.

[9] A. Shokoufandeh, S. Dickinson, K. Siddiqi, and S. Zucker. Indexing using a spectral encoding of topological structure. In *Proceedings, IEEE CVPR*, pages 491–497, Fort Collins, CO, June 1999.

[10] K. Siddiqi, A. Shokoufandeh, S. Dickinson, and S. Zucker. Shock graphs and shape matching. *International Journal of Computer Vision*, 30:1–24, 1999.

[11] Kaleem Siddiqi and Benjamin B. Kimia. A shock grammar for recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 507–513, 1996.