Generic Model Abstraction from Examples

Yakov Keselman, *Member*, *IEEE*, and Sven Dickinson, *Member*, *IEEE*

Abstract—The recognition community has typically avoided bridging the representational gap between traditional, low-level image features and generic models. Instead, the gap has been artificially eliminated by either bringing the image closer to the models using simple scenes containing idealized, textureless objects or by bringing the models closer to the images using 3D CAD model templates or 2D appearance model templates. In this paper, we attempt to bridge the representational gap for the domain of model acquisition. Specifically, we address the problem of automatically acquiring a generic 2D view-based class model from a set of images, each containing an exemplar object belonging to that class. We introduce a novel graph-theoretical formulation of the problem in which we search for the *lowest common abstraction* among a set of lattices, each representing the space of all possible region groupings in a region adjacency graph representation of an input image. The problem is intractable and we present a shortest path-based approximation algorithm to yield an efficient solution. We demonstrate the approach on real imagery.

Index Terms—Image abstraction, automatic model acquisition, learning from examples, shape description, object recognition, graph algorithms.

1 Introduction

1.1 Motivation

In the object recognition community, object representations have spanned a continuum ranging from prototypical models (often called class-based or generic models) to exemplar-based models (often called template-based or appearance-based models). Those advocating prototypical models address the task of recognizing novel (never before seen) exemplars from known classes whose definitions strive to be invariant to changes in surface texture, color, part articulation, and minor deformation in shape. Those advocating exemplar models address the very different task of recognizing particular instances of objects, such as JFK's face or a can of Coke. In a completely orthogonal direction, prototypical models can be object-centered or viewer-centered (or both [14], [13]), provided that the 3D or 2D features that comprise the model satisfy the above invariance goals. Similarly, exemplar models can be object-centered, specifying the exact 3D geometry of an object, e.g., a rigid CAD model "template," or viewer-centered, specifying the exact appearance of an object.

Interestingly, the evolution of object recognition over the past 30 years has followed a path from prototypical models to exemplar models, as illustrated in Fig. 1. Beginning in the 1970s, vision researchers aimed for prototypical vision systems using complex volumetric parts such as generalized cylinders (e.g., [5]) and, later, superquadrics (e.g., [40]) and geons (e.g., [4]). The main challenge to these early systems was the *representational gap* that existed between the low-level features that could be reliably extracted and the abstract nature of the model components. Rather than addressing this representational gap, the community effectively eliminated it by bringing the images closer to the models. This was accomplished by removing object surface markings and structural detail, controlling lighting conditions, and reducing scene clutter. Edges in the image could then be assumed to map directly to the limbs and surface discontinuities of high-order volumetric parts making up the models. The results left many unsatisfied as the images and objects were often contrived and the resulting systems were unable to deal with real objects imaged under real conditions.

The 1980s ushered in 3D models that captured the exact shape of the object. Such models, often in the form of CAD models, were effectively 3D templates, e.g., [22], [32]. Provided that such a model could be acquired for a real object, the community now found that it could build object recognition systems that could begin to recognize real (albeit restricted) objects. This time, the representational gap was eliminated by bringing the model closer to the imaged object, requiring the model to capture the exact geometry of the object. Moreover, since the presence of texture and surface markings seriously affected the computational complexity of these systems, they too selected objects which were texture-free—objects for which a salient image edge discontinuity mapped to a polyhedral edge. Again, there was dissatisfaction as the resulting systems were unable to recognize complex objects with complex surface markings.

Recently, beginning in the 1990s, appearance models have replaced CAD models and, for the first time, recognition systems were constructed that could recognize arbitrarily complex objects, e.g., [50], [36], [31]). Storing a dense set of images of the object from all possible viewpoints, the appearance models were not limited by object geometric complexity, texture, or surface markings. In this case, the representational gap was eliminated by bringing the models all the way down to the image. The resulting systems could therefore recognize only exemplar objects—

- *Y. Keselman is with the School of CTI, DePaul University, 243 S. Wabash Ave., Chicago, IL 60604. E-mail: ykeselman@cti.depaul.edu.*
- *S. Dickinson is with the Department of Computer Science, University of Toronto, 6 King's College Rd., Rm 283B, Pratt Building, Toronto, Ontario, Canada M5S 3G4. E-mail: sven@cs.toronto.edu.*

Manuscript received 05 Jan. 2004; revised 16 Nov. 2004; accepted 15 Dec. 2004; published online 12 May 2005.
Recommended for acceptance by M. Basu.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org and reference IEEECS Log Number TPAMISI-0009-0104.

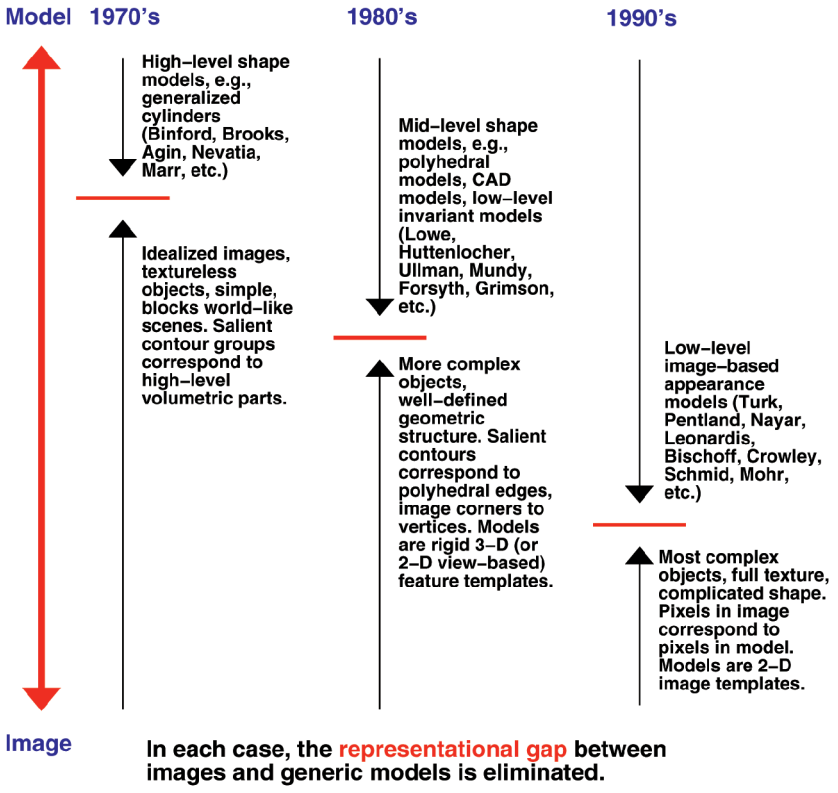

Fig. 1. Evolution of object recognition.

specific objects that had been seen at training time. Despite a number of serious limitations of early approaches to appearance–based recognition, including difficulties in dealing with background clutter, occlusion, object translation, rotation, and scaling, the approach has gained tremendous popularity.

It is important to note that in bringing the model closer to the image, the appearance-based and CAD-based approaches have altered the problem definition from generic to exemplar object recognition. The systems developed in the 70s cannot be compared to those developed today, for their target domains are different. We must acknowledge the efficacy of appearance-based recognition systems for exemplar recognition; provided that the above limitations can be overcome, this technique may emerge as the best method for recognizing exemplars. However, it is important to acknowledge that the prototypical recognition problem is an important one and, despite the vision community's movement toward appearance-based methods, the hypothesis that these (or their analogous 3D template-based) methods can scale up to perform prototypical object recognition is dubious. What, then, has led us away from the important problem of generic object recognition?

Over the past 30 years, the three approaches to eliminating the representational gap (shown in Fig. 1) have been driven by the same limiting assumption: There exists a one-to-one correspondence between a "salient" feature in the image (e.g., a long, high-contrast line or curve, a well-defined homogeneous region, a corner or curvature discontinuity, or, in the case of an appearance-based model, the values of a set of image pixels, possibly restricted to the vicinity of an interest point) and a feature in the model. In the case of generic object recognition, this assumption is fundamentally flawed, for saliency in the image does *not* imply saliency in the model. Under this assumption, object recognition will continue to be exemplar-based and generic recognition will continue to be rather contrived.

To make real progress on the problem of generic object recognition, we must address the representational gap outlined in Fig. 1. Not only must we continue to push the technologies of segmentation and perceptual grouping, we must be able to generate image *abstractions* that may not exist explicitly in the image, but which capture the salient, invariant shape properties of a generic model.[1] We argue that, for the purpose of generic or prototypical object (or, more specifically, shape) recognition, the process of image abstraction must recover a set of "meta-regions" that map to the coarse surfaces (or volumetric primitives) on a prototypical model.

In light of this goal, the difficult challenge of image abstraction can therefore be cast as a twofold problem, as shown in Fig. 2. First, we seek a (compile-time) method for automatically acquiring a generic model from a set of images (containing exemplars belonging to a known class) that bridges the representational gap between the output of an image segmentation module and the "parts" of a generic model. Although, in Fig. 2, this is exemplified by a generic, view-based model (a coffee cup), those advocating

1. Although a number of approaches extract regions or perceptual groups as input to a recognition system, they typically assume that corresponding regions or groups exist on the model.

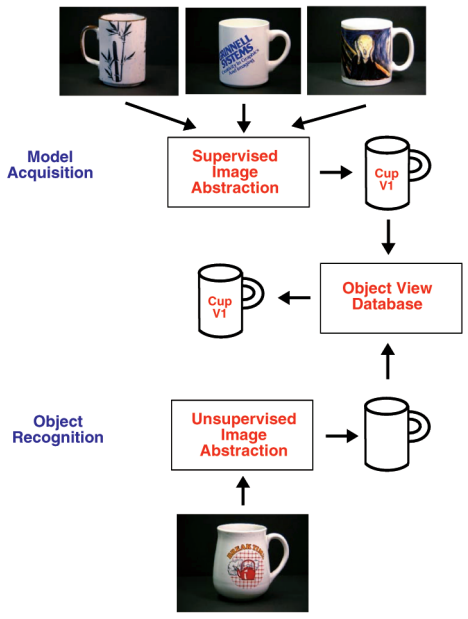

Fig. 2. The role of image abstraction in both model acquisition and object recognition.

object-centered models could easily map the parts of this view into a set of object-centered components, e.g., [14], [13], leading to an object-centered model. Although we make no claim as to the final form of the model (2D or 3D), we believe that a parts-based approach to viewer-centered representations can better accommodate the intractable complexity associated with "whole object" views of complex, articulating objects [14].

Next, from an image of a real exemplar, we seek a (runtime or recognition-time) method that will recover a high-level "abstraction" that contains the coarse features that make up some model. In this paper, we present an approach to the first problem, that of generic model acquisition from a set of exemplars belonging to a known class.[2] The second, and more difficult, problem of generic object recognition is work in progress and is not reported here.[3] Generic object recognition is an essential task whose lack of solution confines object recognition to the laboratory, where such conditions as lighting, clutter, occlusion, and object domain can be tightly controlled. Granted, the generic object recognition torch has continued to burn, albeit dimly, with a number of researchers committed to the problem, e.g., [45], [56], [1], [48], [39], to name just a few. We believe the time has come for the pendulum to swing back toward solving the problem of generic, prototypical, or class-based modeling and recognition. In fact, such a trend is beginning, as witnessed by the movement away from image-based appearance models to robust, interest point features, e.g., [44], [33], [6], and shape contexts (point features with an even larger, albeit qualitative, neighborhood description), e.g., [2].

2. Preliminary algorithms and results have been reported in [27], [26]. This paper expands on these earlier versions, providing additional details on the algorithms and including many new experiments.

3. A discussion of the recognition problem, along with our proposed approaches, appears in Section 6.

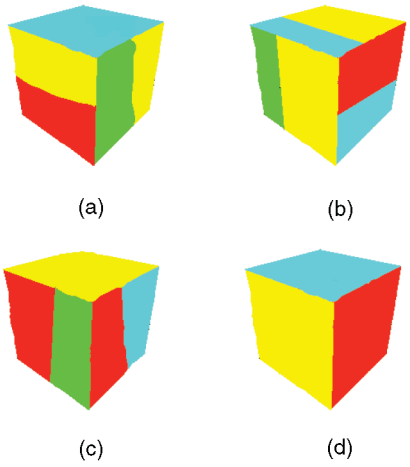

Fig. 3. Illustrative example of generic model acquisition: (a), (b), and (c) input examples belonging to a single known class; (d) generic model abstracted from examples.

1.2 An Illustrative Example

Assume that we are presented with a collection of images such that each image contains a single exemplar, all exemplars belong to a single known class, and that the viewpoint, with respect to the exemplar in each image, is similar. Figs. 3a, 3b, and 3c illustrate a simple example in which three different images, each containing a block in a similar orientation, are presented to the system (we will return to this example throughout the paper to illustrate the various steps in our algorithm). Our task is to find the common structure in these images under the assumption that structure that is common across many exemplars of a known class must be definitive of that class. Fig. 3d illustrates the class "abstraction" that is derived from the input examples. In this case, the domain of input examples is rich enough to "intersect out" irrelevant structure (or appearance) of the block. However, had many or all of the exemplars had vertical stripes, for example, the approach would be expected to include vertical stripes in that view of the abstracted model.

Any discussion of model acquisition must be grounded in image features. In our case, each input image will be region-segmented to yield a region adjacency graph.[4] Similarly, the output of the model acquisition process will yield a region adjacency graph containing the "meta-regions" that define a particular view of the generic model.[5] Other views of the exemplars would similarly yield other views of the generic model. The integration of these views into an optimal partitioning of the viewing sphere or the recovery of 3D parts from these views is beyond the scope of this paper. For now, the result will be a collection of 2D views that describe a generic 3D object. This collection would then be added to the view-based object database used at recognition time.

4. We advocate region segmentation for several reasons. First, surfaces on the object naturally correspond to regions in the image. Second, region shape similarity (a critical component of our algorithm) can be effectively computed using the approach (our previous work) described in [48]. Third, a number of efficient and robust region segmentation algorithms have been recently developed [18], [47], [7].

5. Unlike the regions in the input exemplar region adjacency graphs, the regions of the model region adjacency graphs are salient and can be mapped into meaningful higher-level surface or volumetric primitives.

1.3 Related Work

Automatic model acquisition from images has long been associated with object recognition systems. One of the advantages of appearance-based modeling techniques, e.g., [36], is that no segmentation, grouping, or abstraction is necessary to acquire a model. An object is simply placed on a turntable in front of a camera, the viewing sphere is sampled at an appropriate resolution, and the resulting images (or some clever representation thereof) are stored in a database. Others have sought increased illumination, viewpoint, or occlusion-invariance by extracting local features as opposed to using raw pixel values, e.g., [42], [33], [6], [44], [37].

Appearance-based models (both global and local) are very exemplar-specific due to the specificity of their underlying features, yet restricted categorization is sometimes possible when classes include appearance-based features that are invariant to within-class variability. For example, Mohan et al. [35] have reported promising results in learning component-based, appearance-based models of humans suitable for detection in outdoor cluttered scenes. Leibe and Schiele [30] combine both the appearance and shape (contour) of sets of class exemplars to model and categorize isolated objects. Weber et al. [53] have learned categorical models for heads, leaves, and cars, while Fei-Fei et al. [17] have learned a generative probabilistic model in the form of a constellation of scale-invariant image patches belonging to an object class. Although these approaches attempt to learn categorical descriptions, the categorical features are not true abstractions of the input exemplar features, but, rather, consistently appearing exemplar features.

At the other extreme, the recovery of coarse 3D models from image data has met with mixed success. Triggered by Biederman's introduction of *geons* [4] as a set of abstract 3D shape modeling primitives, many researchers sought to develop algorithms for recovering geons or other qualitative volumetric parts from 2D line drawings or, in certain cases, real images (2D and 3D) [3], [14], [13], [11], [10]. In the domain of range images, greater success has been achieved in extracting coarse models. Generic shape primitives, such as restricted generalized cylinders, quadrics, and superquadrics, have few parameters and can be robustly recovered from 3D range data [41], [49], [19], [12]. Provided the range data can be segmented into parts or surfaces, these generic primitives can be used to model the coarse shapes of the parts, effectively abstracting away structural detail. Unlike methods operating on 2D data, these methods are insensitive to perceived structure in the form of surface markings or texture. However, the above models have been either specified by hand or recovered from a single image, rather than abstracted from a set of examples.

In the domain of recovering 2D generic models from image data, most systems (like many of the appearance-based and 3D modeling frameworks described above) simply fit a prototypical model, e.g., a shock graph [48], [46], to a single image exemplar with the hope that the model will be sufficiently flexible to accommodate other exemplars belonging to its class. In the domain of recovering generic models from a set of 2D examples, there has been considerably less work. The seminal work of Winston [54] pioneered learning descriptions of 3D objects from structural descriptions of positively or negatively labeled examples. Nodes and edges of graph-like structures were annotated with shapes of constituent parts and their relations. As some shapes and relations were abstractions and decompositions of others, the resulting descriptions could be organized into a specificity-based hierarchy. In the 2D shape model domain, Connell and Brady [8] modified Winston's *ANALOGY* program to learn structural concepts from positive examples. However, instead of finding a common abstraction when one-to-one feature correspondence was absent among exemplars belonging to a given class, feature disjunctions were learned. Following on this, Ettinger learned hierarchical, structural descriptions from images based on Brady's curvature primal sketch features [16]. The technique was successfully applied to traffic sign recognition.

An algebraic approach to the automatic construction of structural models was proposed by Nishida and Mori [38] for the problem of learning structural descriptions of characters. Skeletonized character strokes are described in terms of primitive curve elements that meet at junctions and a shape class is defined as a set of structures that can be transformed continuously to each other (not unlike the shape classes defined by Sebastian et al. [46]). An inductive learning step recursively merges pairs of classes that satisfy a set of generalization criteria. Xu et al. [55] address the problem of dynamic learning in a content-based retrieval environment. Given a scene graph in the form of a region tree, combinations of regions are compared to a user-defined template, with a matching combination giving rise to a composite node (abstraction) added to the graph and available for subsequent searches. In a dynamic learning environment, the composite nodes are updated as new templates become available.

On the topic of learning hierarchical structured models from examples, Utans [51] incrementally constructs a Bayesian network in which parts in a compositional hierarchy are grouped into conditionally independent substructrues. Nodes at the lowest levels of the hierarchy represent parts in the input data, while nodes higher up represent groupings of component features into substructures. In recent work, Wachsmuth et al. [52] attempt to learn 2D structural hierarchical descriptions from captioned images. A learned translation model between structured shapes and object nouns is used to guide a search through the space of exemplar abstractions to yield a unifying abstraction that, in turn, leads to a stronger translation model.

In the domain of graph algorithms and computer vision, there have been efforts to generate a prototypical graph from a set of examples. For example, Jiang et al. introduced the concepts of the *median graph* and the *set median graph* [23]. The set median of a set of graphs is defined as that graph, drawn from the set, whose sum distance to the other members (graphs) of the set is minimized, while the median, a more general concept, is not constrained to come from the set. Choosing a prototypical graph from a set of graphs is an important problem in view-based object recognition in which a prototypical graph must be chosen to represent a region on the viewing sphere. Jiang et al. [23]

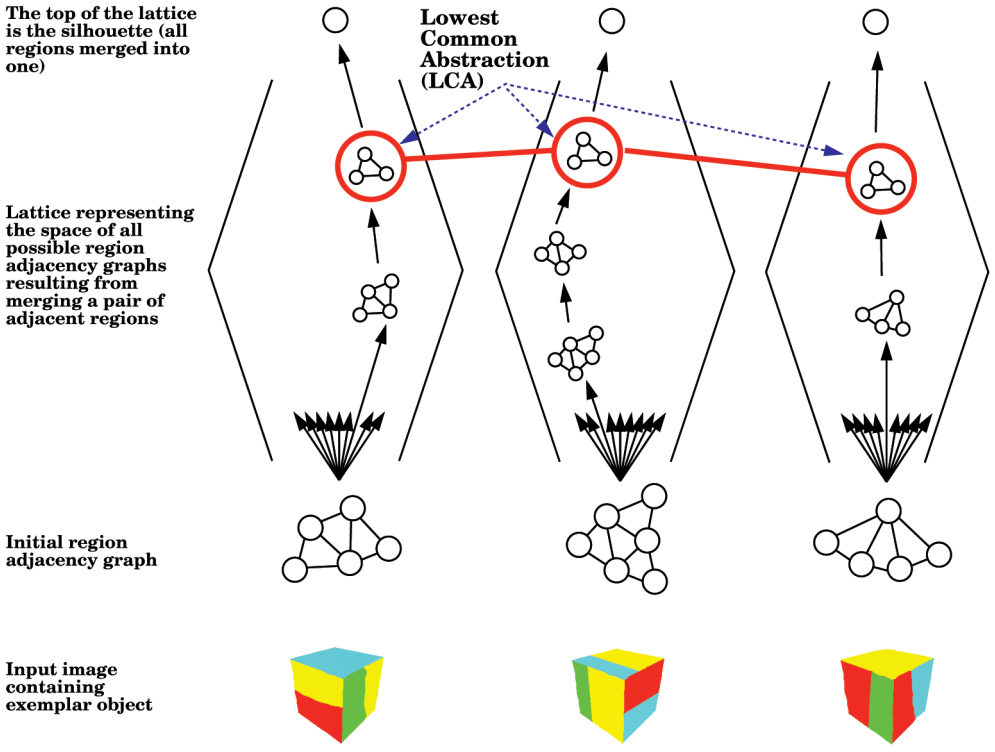

Fig. 4. A Lowest Common Abstraction (LCA) of a set of input exemplars.

proposed a genetic algorithm for computing the generalized median, while Luo et al. have explored the related problem of graph clustering using a spectral embedding of graphs [34]. It is important to note that these approaches assume that graphs belonging to the same class are structurally similar and do not accommodate the many-to-many correspondences common to exemplars belonging to a single class.

1.4 What's Ahead

In the following sections, we begin by presenting a detailed formulation of our problem and conclude that its solution is computationally intractable. Next, we proceed to reformulate our problem by focusing on deriving abstractions from pairs of input images through a top-down procedure that draws on our previous work in generic 2D shape matching. Given a set of pairwise abstractions, we present a novel method for combining them to form an approximation to the solution of our original formulation. We demonstrate the approach by applying it to subsets of images belonging to known classes.

2 Problem Formulation

Returning to Fig. 3, let us now formulate our problem more concretely. As we stated, each input image is processed to form a region adjacency graph (we employ the region segmentation algorithm of Felzenzwalb and Huttenlocher [18]). Let us now consider the region adjacency graph corresponding to one input image. We will assume, for now, that our region adjacency graph represents an oversegmentation of the image. (In Section 6, we will discuss the problem of undersegmentation and how our approach can accommodate it.) The space of all possible region adjacency graphs formed by any sequence of merges of adjacent regions will form a lattice, as shown in Fig. 4. The lattice size is exponential in the number of regions obtained after initial oversegmentation.[6] Kropatsch [29] has studied the problem of structure-preserving graph contraction, leading to techniques that can be used for generating such a lattice.

Each of the input images will yield its own lattice. The bottom node in each lattice will be the original region adjacency graph. In all likelihood, if the exemplars have different shapes (within-class deformations) and/or surface markings, the graphs forming the bottom of their corresponding lattices may bear little or no resemblance to each other. Clearly, similarity between the exemplars cannot be ascertained at this level for there does not exist a one-to-one correspondence between the "salient" features (i.e., regions) in one graph and the salient features in another. On the other hand, the top of each exemplar's lattice, representing a silhouette of the object (where all regions have been merged into one region), carries little information about the salient surfaces of the object.

We can now formulate our problem more precisely, recalling that a lattice consists of a set of nodes with each node corresponding to an entire region adjacency graph. Given N input image exemplars, $E_1, E_2, \ldots, E_N$, let $L_1, L_2, \ldots, L_N$ be their corresponding lattices and, for a given lattice, L_i, let $L_i n_j$ be its constituent nodes, each representing a region adjacency graph, G_{ij}. We define a *common abstraction*, or CA, as a set of nodes (one per

6. Indeed, considering the simple case of a long rectangular strip subdivided into $n+1$ adjacent rectangles, the first pair of mergeable adjacent regions can be selected in n ways, the second in $n-1$, and so on, giving a lattice size of $n!$.

lattice) $L_1 n_{j_1}, L_2 n_{j_2}, \ldots, L_N n_{j_N}$ such that, for any two nodes $L_p n_{j_p}$ and $L_q n_{j_q}$, their corresponding graphs G_{pj_p} and G_{qj_q} are isomorphic. Thus, the root node (whose graph consists of one node representing the silhouette region) of each lattice is a common abstraction. We define a *lowest common abstraction*, or LCA, as a common abstraction whose underlying graph has maximal size (in terms of number of nodes). Given these definitions, our problem can be simply formulated as finding an LCA of N input image exemplars.

Intuitively, we are searching for a node (region segmentation) that is common to every input exemplar's lattice and that retains the maximum amount of structure common to all exemplars. If all the initial exemplars are oversegmented, the desired abstraction can be reached from each initial exemplar by a path corresponding to a sequence of region merges. Unfortunately, the presence of a single, heavily undersegmented exemplar (a single-node silhouette in the extreme case) will drive the LCA toward the trivial silhouette CA. In a later section, we will relax our LCA definition to make it less sensitive to such outliers, effectively allowing an abstraction to be reached from each initial exemplar by a path corresponding to a sequence of region merges *and region splits.*

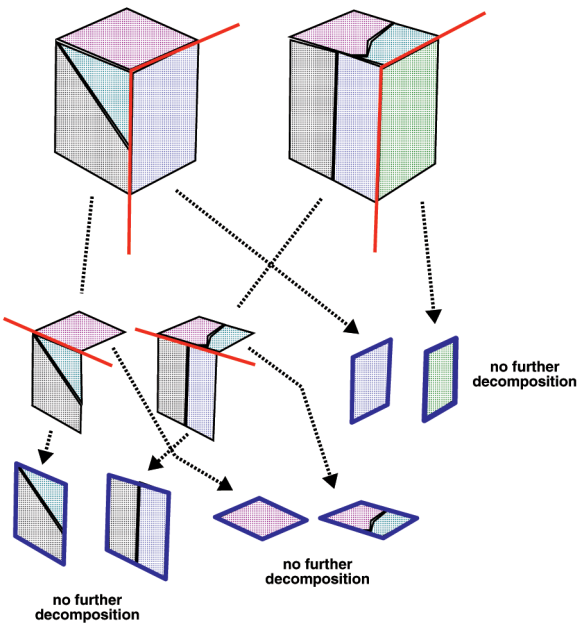

Fig. 5. Finding a lowest common abstraction between two exemplars through a coordinated, recursive decomposition of their silhouettes.

3 The LCA of Two Examples

For the moment, we will focus our attention on finding the LCA of two lattices, while, in the next section, we will accommodate any number of lattices. Since the input lattices are exponential in the number of regions, actually computing the lattices is intractable.[7] Clearly, we need a means for focusing the search for the LCA that avoids significant lattice generation. Our approach will be to restrict the search for the LCA to the *intersection* of the lattices. Typically, the intersection of two lattices is much smaller than either lattice (unless the images are very similar) and leads to a tractable search space. But, how do we generate this new "intersection" search space without enumerating the lattices?

Our solution is to work top-down, beginning with a node known to be in the intersection lattice—the root node, representing a single region (silhouette). If the intersection lattice contains only this one node, i.e., one or both of the region segmented images contain a single region, then the process stops and the LCA is simply the root (silhouette). However, in most cases, the root of each input lattice is derived from an input region adjacency graph containing multiple regions. So, given two silhouettes, each representing the apex of a separate, nontrivial lattice, we have the opportunity to search for a lower abstraction (than the root) common to both lattices. Our approach will be to find a decomposition of each silhouette region into two subregions such that: 1) The shapes of the corresponding subregions are similar and 2) the relations among the corresponding regions are similar. Since there are an infinite number of possible decompositions of a region into two component regions, we will restrict our search to the space of decompositions along region boundaries in the original region adjacency graphs. Note that there may be multiple two-region decompositions that are common to both lattices; each is a member of the intersection set.

The process is illustrated in Fig. 5. The silhouettes of two blocks with their region adjacency graphs overlaid are shown at the top. The best pair of corresponding cuts in the two region adjacency graphs are shown in red such that the two regions to the left of the cuts are similar in shape, the two regions to the right of the cuts are similar in shape, and the relations between the pairs of regions spanning the cuts are similar. The process is repeated recursively, with the left pair of regions decomposed once more, until no further corresponding cuts can be made. The two regions to the right of the top-level cuts are already primitive and cannot be further decomposed. The union of the primitive (blue) regions forms the lowest common abstraction between the two region adjacency graphs. Note that the decomposition is not necessarily unique.

Assuming that we have some means for ranking the matching decompositions (if more than one exists), we pick the best one (the remainder constituting a set of backtracking points) and recursively apply the process to each pair of isomorphic component subregions.[8] The process continues in this fashion, "pushing" its way down the intersection lattice until no further decompositions are found. This lower "fringe" of the search space represents the LCA of the original two lattices. In the following sections, we will formalize this process and examine our approach in detail. An illustration is given in Fig. 6.

3.1 Problem Definition

We define graph H to be an immediate decomposition of graph G if G can be obtained from H by merging two nodes. Let L_1 and L_2 be two lattices and let $G_1 \in L_1$ and $G_2 \in L_2$ be two graphs that are isomorphic. G_1 (or G_2, since they are

7. Thus, any approach that examines significant portions of the lattices, including Frequent Itemset [21], will be computationally infeasible.

8. Each subregion corresponds to the union of a set of regions corresponding to nodes belonging to a connected subgraph of the original region adjacency graph.

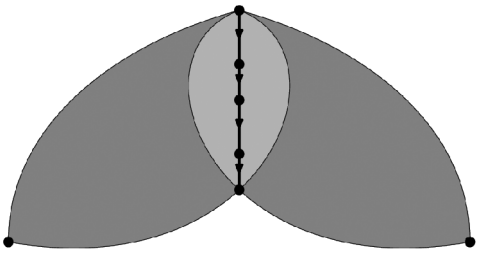

Fig. 6. Abstraction lattices and their intersection. The abstraction lattices of two region adjacency graphs are shown in dark gray. Their intersection, which is significantly smaller, is shown in light gray. To find the lowest common abstraction, we start at the top of the intersection set and work our way down until no further common decompositions are found.

sufficiently similar) is therefore in the intersection of L_1 and L_2. Two graphs, $H_1 \in L_1$ and $H_2 \in L_2$, are common immediate decompositions of G_1 and G_2, respectively, if they are isomorphic immediate decompositions of the respective graphs. Thus, our problem can be formulated as follows: Given a pair of isomorphic graphs G_1 and G_2 in L_1 and L_2, respectively, find a pair of isomorphic immediate decompositions of G_1 and G_2, denoted by $H_1 \in L_1$ and $H_2 \in L_2$, if such a pair exists.

3.2 2D Shape Similarity

Two decompositions (in general, two region adjacency graphs) are isomorphic if their corresponding regions have similar shapes and similar relations. For corresponding regions, it is imperative that we define a similarity metric that accounts for coarse shape similarity. Since the exemplars are all slightly different, so too are the shapes of their abstracted regions. To compute the coarse shape distance between two regions, we draw on our previous approach to generic 2D object recognition that is efficient and robust to occlusion and noise [48]. Specifically, a silhouette (or region boundary, in our case) is decomposed into a set of qualitative parts based on a coloring of the shocks (singularities) of a curve evolution process acting on simple closed curves in the plane [28]. Intuitively, the taxonomy of shocks consists of four distinct types: the radius function along the medial axis varies monotonically at a 1, achieves a strict local minimum at a 2, is constant at a 3, and achieves a strict local maximum at a 4. We have abstracted this system of shocks into a *shock graph*, where vertices are labeled by their shock types and the shock formation times direct the edges (see Fig. 7). The space of such shock graphs is completely characterized by a small number of rules which, in turn, permits the reduction of each graph to a *unique rooted tree*. In prior work, we developed an algorithm for matching two shock trees based on both topological structure and geometric structure [48].

For relational (or arc) similarity, we must check the relational constraints imposed on pairs of corresponding regions. Such constraints can take the form of relative size, relative orientation, or degree of boundary sharing. We implicitly check the consistency of these pairwise constraints by computing the shape distance (using the same distance function referred to above) between the union of the two regions forming one pair (i.e., the union of a region and its neighbor defined by the arc) and the union of the two regions forming the other, as shown in Fig. 8. If the

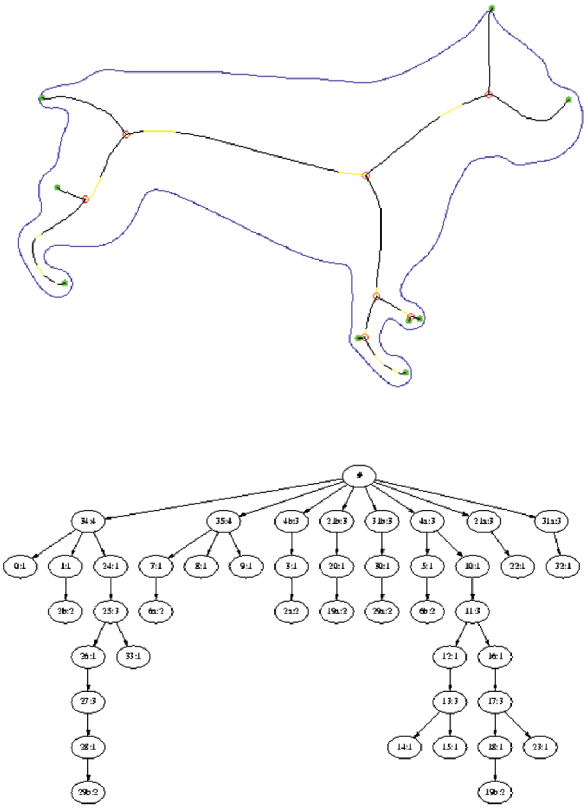

Fig. 7. An illustrative example of a silhouette (top) and its shock graph (bottom).

constraints are satisfied, the distance will be small. In our recursive decomposition framework, the relation between two regions is accounted for by the fact that the parents (i.e., the unions) have already been successfully matched. If the corresponding components match, their relational constraints are therefore satisfied by definition.

3.3 A Shortest Path Formulation

The decomposition of a region into two subregions defines a cut in the original region adjacency subgraph defining the region. Unfortunately, the number of possible two-region decompositions for a given region may be large, particularly for nodes higher in the lattice.[9] One way we can reduce the complexity is to restrict our search for cuts that span two points on the region's boundary, i.e., cuts that don't yield regions with "holes."[10] Despite this restriction, the complexity is still prohibitive and we need to take further measures to simplify our formulation.

We begin by transforming our two region adjacency graphs into their *boundary segment graphs*, as shown in Fig. 9. A boundary segment graph of a region adjacency graph has internal (i.e., common to two original regions) boundary

9. Consider, for example, a checkerboard image and its corresponding region adjacency graph. The root will be a single square region, but there will be *many* decompositions of this square region into two component regions because there are many ways the original checkerboard image can be divided into two along region boundaries. For a checkerboard graph with $(n+1)^2$ vertices, the number of monotonic paths from the upper left corner to the lower right corner is equal to the number of binary sequences of length n, which is exponential.

10. This assumes that a "hole" in a region does not correspond to a salient model surface.

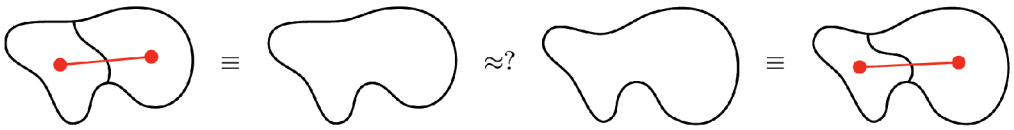

Fig. 8. Checking the relation between two regions. The relation is said to be satisfied if the corresponding regions match *and* their respective unions match.

segments as its nodes and an edge from boundary segment b_1 to b_2 if b_1 and b_2 share an endpoint. Note that, if region A lies entirely within region B, the single boundary segment that is common to A and B is not included in the boundary segment graph. This is due to the fact that we are looking for cuts that yield regions without holes. Note also that nodes in this subgraph can potentially be connected by multiple edges, corresponding to multiple boundary fragments between adjacent regions. This case does not present any difficulties for subsequent stages of our approach.

The transformation to the boundary segment graph allows us to reformulate the search for corresponding cuts in two region adjacency graphs as a search for corresponding paths in their boundary segment graphs.[11] However, this has not affected the complexity of our problem, as there could be an exponential number of paths in each boundary segment graph (recall our checkerboard example). Rather than enumerating the paths in each boundary segment graph and then enumerating all pairs, we will cast our problem as a search for the shortest path in a *product graph* of the two boundary segment graphs.

The product graph $G = G_1 \times G_2 = (V, E)$ of graphs $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$ is defined as follows:

$$V = \{(v_1, v_2) : v_1 \in V_1,\ v_2 \in V_2\} = V_1 \times V_2,$$

$$E = \begin{array}{ll} \{((u_1,u_2),(v_1,v_2)) : & (u_1,v_1) \in E_1, (u_2,v_2) \in E_2\ \} \cup \\ \{((v_1,u_2),(v_1,v_2)) : & v_1 \in V_1, (u_2,v_2) \in E_2\ \} \cup \\ \{((u_1,v_2),(v_1,v_2)) : & (u_1,v_1) \in E_1, \quad v_2 \in V_2\}. \end{array}$$

The node set of the product graph is the product of the node sets of the initial graphs. To define the edge set of the product graph correctly, notice that the simple product of the edge sets (the first term in the union) may result in a disconnected graph. The other two terms of the union (which can be viewed as the product of the node set of one graph with the edge set of the other graph) ensure that the product of two connected graphs will be connected. The essential property of the product graph that we will exploit is that a simple path $(u_1,v_1) \to (u_2,v_2) \to \cdots \to (u_n,v_n)$ in the product graph corresponds to two sequences of nodes in the initial graphs, $u_1 \to u_2 \to \cdots \to u_n$ and $v_1 \to v_2 \to \cdots \to v_n$ which, after the elimination of successive repeated nodes, will result in two paths (whose lengths may be different) in the initial graphs, as shown in Fig. 10.

Each path in our product graph corresponds to a pair of possible cuts in two regions. Consequently, our goal is to find a path that yields the best matching subregions and relations. Here is where we face a problem. This objective function can only be evaluated once a complete path is found since only a complete path defines a pair of closed regions whose shapes and relations can be matched. However, a given edge in the product graph represents a pair of corresponding boundary fragments from two regions. Globally, we are comparing regions, while, locally, we are comparing contours. How then do we define local edge weights and a local objective function so that a shortest path algorithm will yield an approximation to the global optimum?

Let us begin with the edge weights. If we align the two regions (from which we seek corresponding cuts) through our region matching algorithm [48], then we can assume that both external and internal boundary segments are approximately aligned. Under this assumption, edge weights in the product graph can be chosen to reflect the shape similarity of their component boundary segments. In our implementation, to compute edge weights, we employ a simple Hausdorff-like distance between the two boundary segments, yielding a local approximation to the global similarity of the regions.

As a result, in the product of two boundary segment graphs, smaller edge weights correspond to pairs of more similar boundary segments. This leads to a number of very natural choices for an objective function if we interpret edge weights as edge lengths. The total path length, $tl(p) = \sum_{p_i \in p} l(p_i)$, is a well-studied objective function [9]. Fast algorithms for generating successive (i.e., next optimal) shortest and simple shortest paths are given in [15], [25]. However, the above objective function tends to prefer shorter paths over longer ones, assuming path edges are of equal average length. For our problem, smaller paths will result in smaller regions being cut off, which is contrary to our goal of finding the lowest common abstraction.[12]

To overcome this problem, we turn to a different objective function that measures the maximum edge weight on a path, $ml(p) = \max_{p_i \in p} l(p_i)$. A well-known modification[13] of Dijkstra's algorithm [9] finds paths of minimal maximum edge weight (minmax paths) between a chosen node and all other graph nodes and has the same complexity, $O(|E| + |V| \log |V|)$, as the original algorithm. However, efficient algorithms for finding successive (next optimal) minmax paths are not readily available. Leaving development of such an algorithm for the future, we will employ a mixed strategy. Namely, we find pairs of nodes providing near-optimal values of the minmax objective function and, along with the minmax path between the nodes, we also generate several successive shortest paths between them. For this, we use Eppstein's algorithm [15], which generates k successive shortest paths between a chosen pair of nodes in $O(|E| + |V| \log |V| + k \log k)$ time. The mixed strategy, whose overall complexity is $O(|V|(|E| + |V| \log |V|))$ for small k, has proven to be effective in empirical testing.

11. Thus, our boundary segment graph can be thought of as a weaker version of the dual graph [20].

12. A small region is unlikely to be common to many input exemplars.

13. Instead of summing up edge weights when determining the distance to a node, it takes their maximum.

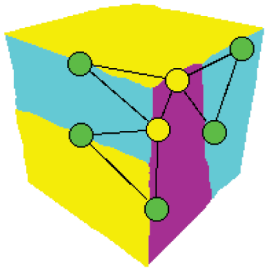

Fig. 9. A region decomposition and its boundary segment graph. Green nodes correspond to boundary segments that touch the background. Yellow nodes correspond to boundary segments that do not touch the background. A cut in the region decomposition corresponds to a path between a pair of green nodes.

3.4 Algorithm

Having defined the edge weights and objective function, we can now summarize our algorithm for finding the "best" common decomposition of two abstraction nodes, as shown in Algorithm 1. This algorithm, in turn, is embedded in our solution to the problem of finding the LCA of two examples, which computes an approximation to the intersection of their respective lattices in a top-down manner. Beginning with the two root nodes (the sole member of the initialized intersection set), we recursively seek the "best" common decomposition of these nodes and add it to the intersection set. The process is recursively applied to each common decomposition (i.e., member of the intersection set) until no further common decompositions are found. The resulting set of "lowest" common decompositions represents the LCA of the two lattices. The description is formalized in Algorithm 2.

Algorithm 1. A generic algorithm for finding a common decomposition.

1. Let A_1, A_2 be subgraphs of the original region adjacency graphs that correspond to isomorphic vertices of the abstraction graphs.
2. Let G_1, G_2 be boundary segment graphs of A_1, A_2.
3. Form the product graph $G = G_1 \times G_2$ as described above.
4. Choose an objective function f (see text for discussion), compute edge weights w_i (see text for discussion), and select a threshold $\varepsilon > 0$.
5. Let P_f be the optimal path with respect to $(f, \{w_i\})$ with value $F(P_f)$.
6. Let $P = P_f$
7. **while** $|f(P) - f(P_f)| < \varepsilon$ **do**
8. Let P_1 and P_2 be the paths in G_1, G_2 corresponding to P.
9. Let (V_1, W_1) and (V_2, W_2) be the resulting cuts in A_1, A_2
10. **if** region V_1 is similar to region V_2 and region W_1 is similar to region W_2 and arcs (V_1, U_1^i), (V_2, U_2^i) are similar for all isomorphic neighbors U_1^i, U_2^i of V_1, V_2, respectively, and arcs (W_1, U_1^i), (W_2, U_2^i) are similar for all isomorphic neighbors U_1^i, U_2^i of W_1, W_2, respectively, **then**
11. **output** decompositions (V_1, W_1) and (V_2, W_2).
12. **return**
13. **end if**
14. Let P be the next optimal path with respect to $(f, \{w_i\})$.
15. **end while**
16. **output** "no nontrivial decomposition is found."

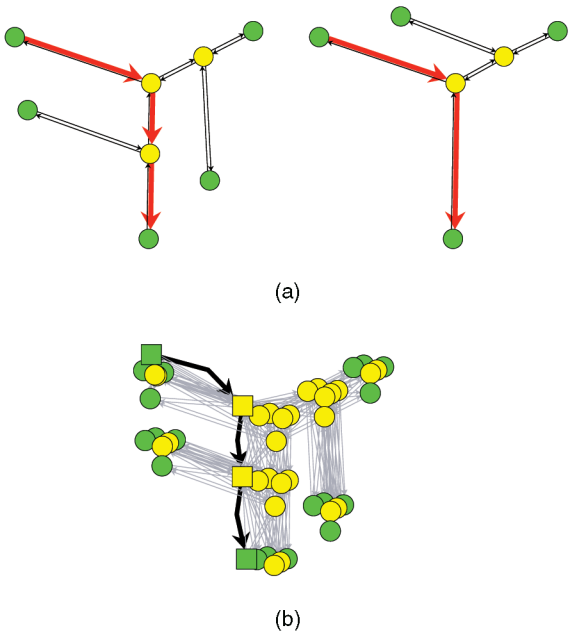

Fig. 10. The product graph. The product graph of the two graphs in (a) is shown in (b). The pair of red paths shown in the original graphs corresponds to the black path shown in the product graph.

Algorithm 2. Finding the maximal common abstraction of two region adjacency graphs.

1. Let A_1, A_2 be the initial region adjacency graphs.
2. Let G_1, G_2 denote abstraction graphs belonging to abstraction lattices, L_1 and L_2, respectively.
3. Let G_1^0, G_2^0 be the topmost nodes of the lattices.
4. Let $G_1 = G_1^0$, $G_2 = G_2^0$.
5. **while** there are unexplored isomorphic nodes $u_1 \in G_1$, $u_2 \in G_2$ **do**
6. Let U_1 and U_2 be the corresponding subgraphs of A_1, A_2.
7. **if** there is a *common decomposition* $U_1 = V_1 \cup W_1$ and $U_2 = V_2 \cup W_2$ **then**
8. Split the nodes $u_1 \in G_1$, $u_2 \in G_2$ by forming the *decomposition graphs* $H_1 = (G_1 - \{u_1\}) \cup \{v_1, w_1\}$, $H_2 = (G_2 - \{u_2\}) \cup \{v_2, w_2\}$ with edges established using A_1, A_2.
9. Let $G_1 = H_1$, $G_2 = H_2$.
10. **else**
11. Mark u_1 and u_2 as explored.
12. **end if**
13. **end while**
14. **output** G_1, G_2.

4 The LCA of Multiple Examples

So far, we have addressed only the problem of finding the LCA of two examples. How then can we extend our approach to find the LCA of multiple examples? Furthermore, when moving toward multiple examples, how do we prevent a "noisy" example, such as a single, heavily undersegmented silhouette from derailing the search for a

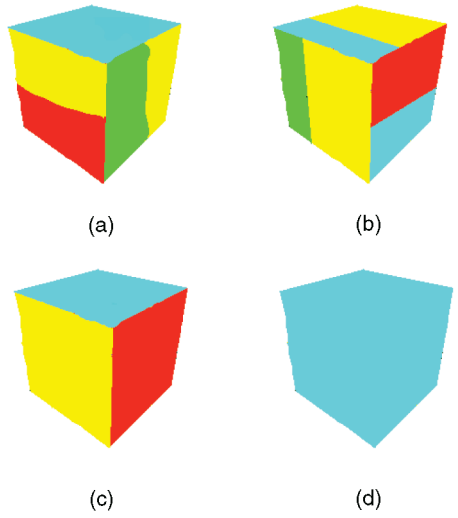

Fig. 11. The straightforward computation of a Lowest Common Abstraction of exemplars (a)-(d) gives the exemplar in (d). However, (c) is a Lowest Common Abstraction of exemplars (a)-(c) and, therefore, is more representative.

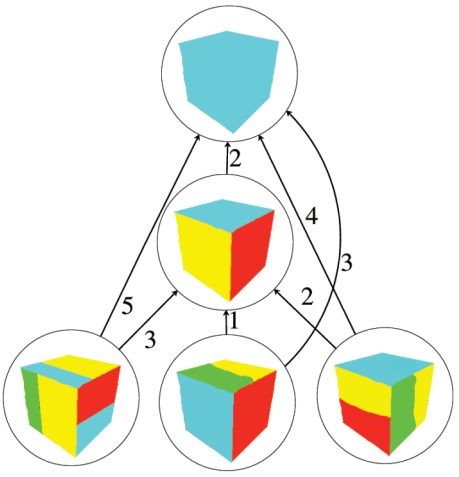

Fig. 12. Embedding region adjacency graphs and their pairwise LCAs in a weighted directed acyclic graph. The three nodes at the bottom, as well as the undersegmented "outlier" at the top, are the four input exemplars. Although the true LCA is the top node, the center node is the median, as its distance sum value is $3+1+2+2=8$, while the sum is $5+3+4+0=12$ for the topmost node.

meaningful LCA? To illustrate this effect, consider the inputs shown in Figs. 11a, 11b, 11c, and 11d. If the definition of the pairwise LCA is directly generalized, thus requiring the search for an element common to all abstraction lattices, the correct answer will be the input Fig. 11d. However, much useful structure is apparent in inputs Figs. 11a, 11b, and 11c; input Fig. 11d can be considered to be an outlier.

To extend our two-exemplar LCA solution to a robust, multi-exemplar solution, we begin with two important observations. First, the LCA of two exemplars lies in the intersection of their abstraction lattices. Thus, both exemplar region adjacency graphs can be transformed into their LCA by means of sequences of region merges. Second, the total number of merges required to transform the graphs into their LCA is minimal among all elements of the intersection lattice, i.e., the LCA lies at the lower fringe of the lattice.

Our solution begins by constructing an approximation to the intersection lattice of multiple exemplars. Consider the closure of the set of the original region adjacency graphs under the operation of taking pairwise LCAs. In other words, starting with the initial region adjacency graphs, we find their pairwise LCAs, then find pairwise LCAs of the resulting abstraction graphs, and so on (note that duplicate graphs are removed). We take all graphs, original and LCA, to be nodes of a new *closure* graph. If graph H was obtained as the LCA of graphs G_1 and G_2, then directed arcs go from nodes corresponding to G_1, G_2 to the node corresponding to H in the closure graph.

Next, we will relax the first property above to accommodate "outlier" exemplars, such as undersegmented input silhouettes. Specifically, we will not enforce that the LCA of multiple exemplars lie in the intersection set of *all* input exemplars. Rather, we will choose a node in our approximate intersection lattice that represents a "low abstraction" for many (but not necessarily all) input exemplars. More formally, we will define the LCA of a set of exemplar region adjacency graphs to be that element in the intersection of two or more abstraction lattices that minimizes the total number of edit operations (merges or splits) required to obtain the element from *all* the given exemplars. If a node in the intersection lattice lies along the lower fringe with respect to a number of input exemplars, then its sum distance to all exemplars is small. Conversely, the sum distance between the silhouette outlier (in fact, the true LCA) and the other input exemplars will be large, eliminating that node from contention.

Note that a graph may not be directly linked to all of its abstractions in the closure graph. However, if H is an abstraction of G, then there is a directed path between the nodes corresponding to G and H. Thus, any abstraction is reachable from any of its decompositions by a directed path. Such a path may move upward from an exemplar through zero or more successive region merges to a node in the closure graph, then downward through zero or more successive region splits to reach the abstraction. Each edge in the closure graph is assigned a weight equal to the merge edit distance that takes the decomposition to the abstraction. The edit distance is simply the difference between the numbers of nodes in the decomposition graph and the abstraction graph. As a result, we obtain a weighted directed acyclic graph. An example of such a graph, whose edges are shown directed from region adjacency graphs to their LCAs, is given in Fig. 12.

Given such a graph, the robust LCA of *all* inputs will be that node that minimizes the sum of shortest path distances from the initial adjacency graphs. In other words, we are looking for the "median" of the graph as computed by Algorithm 3. Note that the closure graph is an approximation to the intersection lattice. On one hand, it may contain pairwise LCAs which are not contained in the intersection lattice, while, on the other hand, it may not contain nodes in the intersection lattice that are not LCAs. While it will contain the true LCA, our median formulation may lower the LCA fringe below the true LCA.

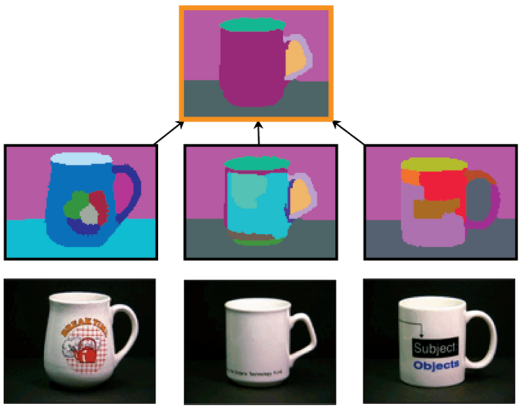

Fig. 13. Computed LCA (orange border) of three examples.

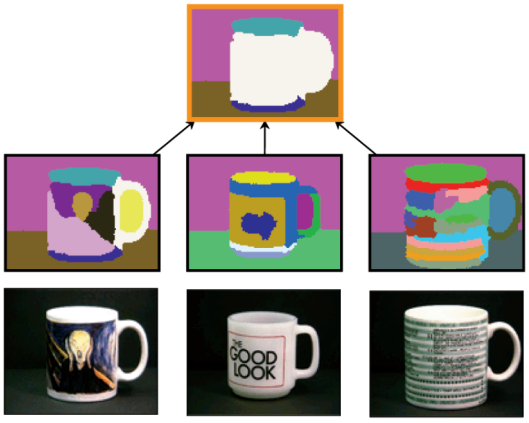

Fig. 14. Computed LCA (orange border) of three examples.

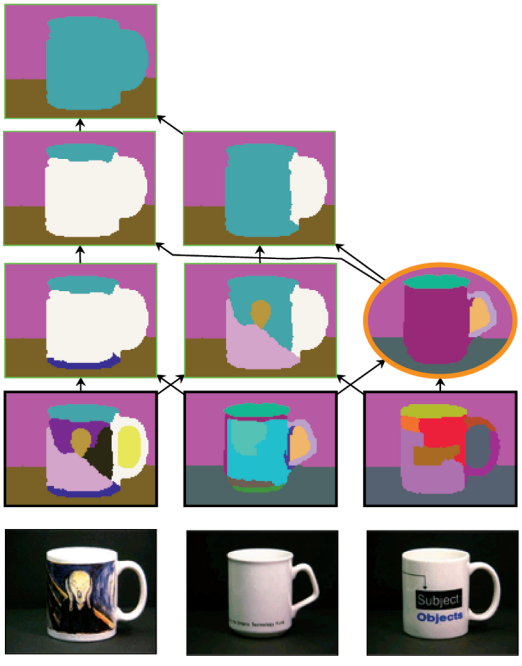

Fig. 15. Computed LCA (orange border) of three examples. Each nonleaf node in the closure graph is the LCA of the two nodes from which edges are directed to it.

Algorithm 3. Finding the median of the closure graph.
1. Let the *sink node*, s, be the topmost node in the closure graph.
2. Solve the "many-to-one" directed shortest path problem on the graph with the source nodes being the original adjacency graphs and with the specified edge weights. Find the distance sum, DS(s), for the sink node.
3. Similarly, find distance sums, $DS(s_i)$, for all unexplored $s_i \in N(s)$.
4. **if** $min_i(DS(s_i)) \geq DS(s)$ **then**
5. **return** s
6. **else**
7. Let $s = \arg\min_i DS(s_i)$.
8. **goto** 2.
9. **end if**

To analyze the complexity of the algorithm, notice that the first step, i.e., finding the distance sum to the topmost node, can be performed in linear time in the graph size since the closure graph is a directed acyclic graph and the single source shortest path problem in such graphs can be solved in $O(|V| + |E|)$ time [9]. Since the algorithm can potentially examine a constant fraction of the graph nodes (consider the case of a line graph), the total running time can be as high as $O(|V|(|V| + |E|))$. The average case complexity will depend on the particular distribution of the initial data and is beyond the scope of this paper. In practice, the algorithm converges in a few iterations.

5 Experiments

In this section, we apply our approach to three different object domains. The domains of coffee cups, books, and dispenser bottles were chosen since they can be effectively modeled as collections of surfaces in 3D, different exemplars belonging to a class exhibit minor within-class shape deformation and the different exemplars have significantly different low-level appearance in terms of color, texture, and surface markings. In Fig. 13 and Fig. 14, we illustrate the results of our approach applied to two sets of three coffee cup images, respectively. In each case, the lower row represents the original images, the next row up represents the input region segmented images (with black borders), while the LCA is shown with an orange border.[14] In each case, the closure graph consists of only four members, with the same pairwise LCA emerging from all input pairs. While, in Fig. 13, the solution captures our intuitive notion of the cup's surfaces, the solution in Fig. 14 merges the regions corresponding to the surfaces defined by the cup's body and handle. A strip along the bottom is present in each exemplar and understandably becomes part of the solution. The same is true of the elliptical region at the top of each exemplar. However, due to region segmentation errors, the larger blue region in the middle cup extends into the handle. Consequently, a cut along its handle (as is possible on the other cups) is not possible for this exemplar, resulting in a "stopping short" of the recursive decomposition at the large white region in the solution (LCA).

In Fig. 15, we again present three exemplars to the system. In this case, the closure graph has many nodes.

14. All region segmented images are the actual outputs of the implementation that accompanied [18]. In particular, smaller regions are sometimes subsumed by larger regions, resulting in nonintuitive segmentations.

Fig. 16. The set of eight cup images (and their region segmentations) used in the experiments.

Unlike Fig. 13 and Fig. 14, in which all pairwise LCAs were equal (leading to a somewhat trivial solution to our search for the global LCA), each pair of input exemplars leads to a different LCA which, in turn, leads to additional LCAs. Continuing this process eventually results in the inclusion of the silhouette in the closure graph. The solution according to our algorithm is again shown in orange and represents an effective model for the cup.

To test our approach more systematically, we applied it to subsets of the cup images shown in Fig. 16. Specifically, we randomly selected four subsets, with three to four images each. A typical result is shown in Fig. 17. Based on the results of this and other experiments, it turned out that all but one segmentation had the four-region LCA as its abstraction, while no other region adjacency graph was an abstraction of more than two exemplars. This means that the LCA of these four cup images is also the LCA of the whole set.

This observation suggests another approach to finding the abstraction of a large number of exemplars: Generate the LCA of a small subset (two to four) of exemplars and check (by finding pairwise LCAs) whether the LCA is an abstraction of many other exemplars. It can be shown that an abstraction of more than half the initial exemplars is also an abstraction of their LCA. Moreover, the largest such "small subset" LCA (found, for example, by exhaustive search among all "small subset" LCAs) will be isomorphic to the one found with the current exhaustive approach. This alternative sample-based approach (similar to robust statistical estimation or to learning from representative samples) will not require the computation of the full closure graph, which is the main bottleneck of our current approach. Its implementation is left for future work.

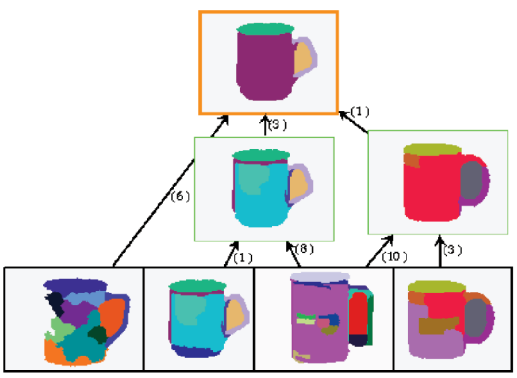

Fig. 17. Computed LCA (orange border) of four examples. Each nonleaf node in the closure graph is the LCA of the nodes from which edges are directed to it. Each edge label indicates the edit distance between the region adjacency graphs corresponding to the two nodes that span the edge. The computed LCA is an abstraction of seven out of eight initial exemplars.

To illustrate the fact that the LCA of the entire set of exemplars can be equal to the LCA of just two exemplars, we turn to the domain of books, a set of eight exemplars of which is shown in Fig. 18; as before, for each book, we include the original image and the segmented version that serves as input to our algorithm. In Fig. 19, the computed LCA (orange border), which is a pairwise LCA of the left two exemplar region adjacency graphs, is in fact the LCA of the entire eight-image set. This is due to the fact that a stripe-like region at the top of the "horizontal" surface patch is present in all but two book images. On the other hand, all other regions on the "horizontal" surface patch of the book model disappear once sufficiently many exemplars are shown to the system. The resulting LCA suggests that many such computer science texts share a horizontal stripe at the top of one of their covers.

To illustrate the fact that the LCA of a relatively large subset (in this case, three of eight) of input exemplars may not necessarily be the LCA of the entire set (of eight), consider the computed LCA of the three book exemplars shown in Fig. 20. Due to the presence of a heavily oversegmented exemplar (the rightmost node), the resulting LCA (orange border) is skewed toward having too many regions. The apparent regularity of such a structure renders it salient. This "skew" can be corrected by either choosing a different subset of images or by adding more exemplars to the subset. To illustrate the effect on the LCA of adding exemplars, consider the domain of dispenser bottles, as shown in Fig. 21. The closure graphs of sets of three and four dispenser bottles are shown in Fig. 22 and Fig. 23, respectively. In Fig. 22, even though the desired three-region LCA of the three exemplars appears in the closure graph (top node), the computed LCA is the undesirable four-region abstraction. This is due to the fact that the accidental alignment of nonsalient structure in the right two exemplars renders the leftmost exemplar an "outlier." The addition of a fourth exemplar overcomes this problem, leading to the desired LCA, as shown in Fig. 23.

In concluding this section, it is worth addressing the issue of background in the model acquisition process. As is evident in the above experiments, all exemplars were imaged against simple, mostly uniform backgrounds (in some cases, the table on which the cup was sitting was segmented from the background). What if each exemplar appeared against a cluttered background? If the cluttered background was consistently seen in many of the images, the algorithm would, of course, not be able to distinguish this regularly occurring structure from that of the object. If the backgrounds were inconsistent and we could assume that the object's silhouettes were the most salient regions in the images, then a pair of large, similar regions could be extracted from a pair of images by an application of the product graph technique as follows. Recalling that a region

Fig. 18. The set of eight book images (and their region segmentations) used in the experiments.

in the image corresponds to a cycle in the boundary segment graph, a pair of similar regions corresponds to a pair of cycles in the boundary segment graphs which, in turn, corresponds to a single "optimal" cycle in the product graph. To find the optimal cycle in the product graph, we can apply the minimum mean cycle algorithm [9]. Once the foreground objects have been separated from their backgrounds, our "standard" procedure can be applied.

6 Conclusions

The quest for generic object recognition hinges on an ability to generate abstract, high-level descriptions of input data. This process is essential not only at runtime for the recognition of objects, but also at compile time for the automatic acquisition of generic object models. In this paper, we address the latter problem—that of generic model acquisition from examples. We have introduced a novel formulation of the problem in which the model is defined as the lowest common abstraction of a number of segmentation lattices, representing a set of input image exemplars. To manage the intractable complexity of this formulation, we focus our search on the intersection of the lattices, reducing complexity by first considering pairs of lattices and later combining these local results to yield an approximation to the global solution.[15]

We have demonstrated the framework on three separate domains (cups, books, and dispenser bottles) in which a generic, view-based model (for the canonical view) is computed from a small set of exemplars. Although these results are encouraging, it should be noted that we have made a very important assumption that region segmentation errors exist in the form of oversegmentation. Although most region segmentation algorithms can be "pushed" toward oversegmentation, undersegmentation cannot be avoided altogether, requiring some means for splitting regions. Unfortunately, for any region, there are an infinite number of ways of splitting the region. Fortunately, our region representation (a shock graph [48]) explicitly encodes a small, finite number of region split points, allowing us to accommodate region splitting within our framework in a tractable way. But where should such splitting occur in the process? One possibility would be to generate additional input exemplars from existing ones by splitting regions in various ways. A more efficient approach would be to incorporate the splitting process into the recursive decomposition of two input exemplars. In the case where one region is primitive and its corresponding region is not, the decomposition would normally terminate. However, if the primitive region can be split (among its finite split possibilities) so that decomposition can continue, the split is retained. We plan to explore region splitting in future work.

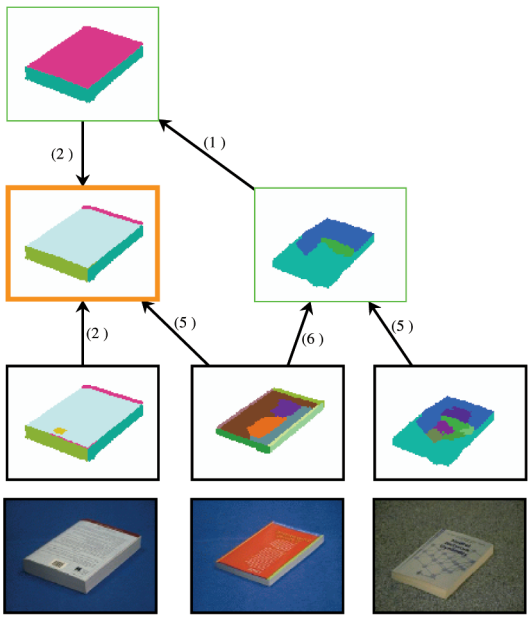

Fig. 19. Computed LCA (orange border) of three examples. The solution is also the LCA of the entire set of eight book images.

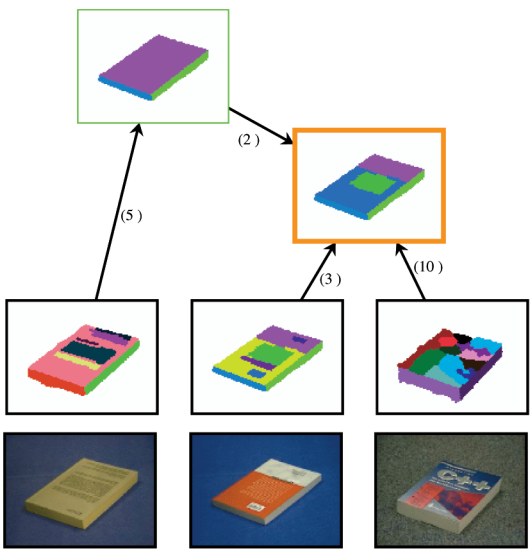

Fig. 20. Computed LCA (orange border) of three examples. The solution, influenced by the heavily oversegmented exemplar (rightmost node), is a decomposition of the desired LCA (top node).

15. Although we solved the original intractable problem only approximately, the resulting approximate solution gives an effective object model that can be used in subsequent object recognition. Our reformulation can be compared with that used in machine learning: The search for the best classifier is often intractable, but the approximately best models are used for the actual classification with good results.

Fig. 21. The set of four dispenser bottle images (and their region segmentations) used in the experiments.

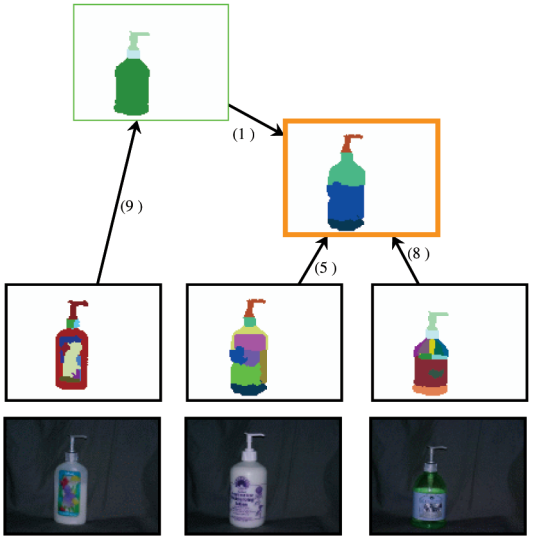

Fig. 22. Computed LCA (orange border) of three examples. Accidental alignment of nonsalient structure in the rightmost two exemplars leads to an undesirable LCA.

Very little world knowledge is currently used to constrain the abstraction process. In certain domains, particular region shapes may be prominent while others are impossible. Any such knowledge, in the form of known region shape or region adjacency constraints, can be used to prune the search space of possible cuts in a region adjacency graph. Perhaps even an object class's known functionality can be used to constrain the abstraction process in terms of its mapping to underlying shape constraints [43]. Object appearance might also be exploited when *general* constraints are known on object color and/or texture.

We have focused on the generic model acquisition problem, leaving the more difficult problem of generic object recognition to future work. Toward the more restricted goal of finding a particular *target* object in an image, our approach to finding pairs of similar regions can be easily adapted as follows (assuming a region oversegmentation of the image): Given a region adjacency graph representing an object model, we first try to find in the image the object's silhouette. This can be accomplished by finding the minimum mean cycle in the product of the model's silhouette graph and the image's boundary segment graph. Having found the best match for the object's silhouette, we compute product graphs and minimum mean cycles at the level of individual regions. If matches are established at all levels, the model is detected in the image. Thus, through an application of our model acquisition technique, target models can be detected in an image.

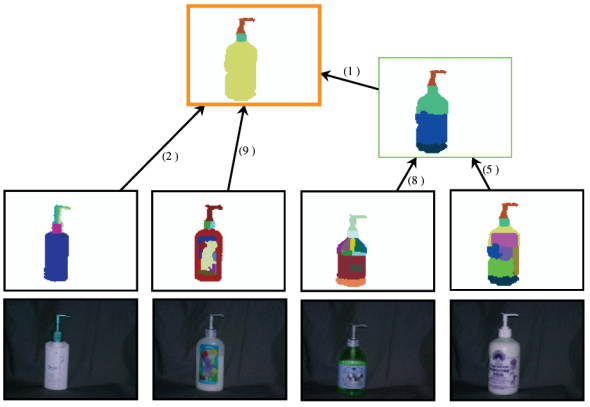

Fig. 23. Computed LCA (orange border) of four examples. The additional exemplar results in the desirable LCA.

Our next major step is the less constrained problem of *unexpected* object recognition of a novel exemplar from our acquired object classes. Our efforts are currently focused on the analysis of the conditions under which two regions are merged. If we can derive a set of rules for the perceptual grouping of regions, we will be able to generate abstractions from images. Given a rich set of training data derived from the model acquisition process (recall that the LCA of two examples yields a path of region merges along with a set of "no-merges"), we are currently applying machine learning methods to uncover these conditions. Combined with our model acquisition procedure, we can close the loop on a system for generic object recognition which addresses a representational gap that has typically been ignored in computer vision.

Acknowledgments

The authors would like to thank Allan Jepson for his insightful comments and feedback on this work, as well as the three reviewers, whose comments have improved the presentation. The authors would also like to gratefully acknowledge the generous financial support of the US Army Research Office, the US National Science Foundation, the Natural Sciences and Engineering Research Council of Canada, and the Province of Ontario (PREA).

References

[1] R. Basri, L. Costa, D. Geiger, and D. Jacobs, "Determining the Similarity of Deformable Shapes," *Proc. Int'l Conf. Computer Vision Workshop Physics-Based Modeling in Computer Vision*, pp. 135-143, 1995.

[2] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509-522, Apr. 2002.

[3] R. Bergevin and M.D. Levine, "Generic Object Recognition: Building and Matching Coarse Descriptions from Line Drawings," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no 1, pp. 19-36, Jan. 1993.

[4] I. Biederman, "Human Image Understanding: Recent Research and a Theory," *Computer Vision, Graphics, and Image Processing*, vol. 32, pp. 29-73, 1985.

[5] R. Brooks, "Model-Based 3-D Interpretations of 2-D Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 5, no. 2, pp. 140-150, 1983.

[6] G. Carneiro and A. Jepson, "Local Phase-Based Features," *Proc. European Conf. Computer Vision*, 2002.

[7] D. Comaniciu and P. Meer, "Robust Analysis of Feature Spaces: Color Image Segmentation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 750-755, 1997.
[8] J. Connell and M. Brady, "Generating and Generalizing Models of Visual Objects," *Artificial Intelligence,* vol. 31, pp. 159-183, 1987.
[9] T.H. Cormen, C.E. Leiserson, and R.L. Rivest, *Introduction to Algorithms,* chapter 25. MIT Press, 1993.
[10] S. Dickinson, H. Christensen, J. Tsotsos, and G. Olofsson, "Active Object Recognition Integrating Attention and Viewpoint Control," *Computer Vision and Image Understanding,* vol. 67, no. 3, pp. 239-260, Sept. 1997.
[11] S. Dickinson and D. Metaxas, "Integrating Qualitative and Quantitative Shape Recovery," *Int'l J. Computer Vision,* vol. 13, no. 3, pp. 1-20, 1994.
[12] S. Dickinson, D. Metaxas, and A. Pentland, "The Role of Model-Based Segmentation in the Recovery of Volumetric Parts from Range Data," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 3, pp. 259-267, Mar. 1997.
[13] S. Dickinson, A. Pentland, and A. Rosenfeld, "From Volumes to Views: An Approach to 3-D Object Recognition," *Computer Graphics, Vision, and Image Processing: Image Understanding,* vol. 55, no. 2, pp. 130-154, 1992.
[14] S. Dickinson, A. Pentland, and A. Rosenfeld, "3-D Shape Recovery Using Distributed Aspect Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 14, no. 2, pp. 174-198, Feb. 1992.
[15] D. Eppstein, "Finding the k Shortest Paths," *SIAM J. Computing,* vol. 28, no. 2, pp. 652-673, 1999.
[16] G. Ettinger, "Large Hierarchical Object Recognition Using Libraries of Parameterized Model Sub-Parts," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 32-41, 1988.
[17] L. Fei-Fei, R. Fergus, and P. Perona, "Learning Generative Visual Models from Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object Categories," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* June 2004.
[18] P. Felzenszwalb and D. Huttenlocher, "Image Segmentation Using Local Variation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 98-104, 1998.
[19] F. Ferrie, J. Lagarde, and P. Whaite, "Darboux Frames, Snakes, and Super-Quadrics: Geometry from the Bottom Up," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 15, no. 8, pp. 771-784, Aug. 1993.
[20] R. Gould, *Graph Theory,* pp. 170-172. Benjamin/Cummings, 1988.
[21] J. Han and M. Kamber, *Data Mining: Concepts and Techniques,* p. 228. Morgan Kaufmann, 2001.
[22] D. Huttenlocher and S. Ullman, "Recognizing Solid Objects by Alignment with an Image," *Int'l J. Computer Vision,* vol. 5, no. 2, pp. 195-212, 1990.
[23] X. Jiang, A. Munger, and H. Bunke, "On Median Graphs: Properties, Algorithms, and Applications," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 10, Oct. 2001.
[24] J.-M. Jolion, "The Deviation of a Set of Strings," *Pattern Analysis and Applications,* vol. 6, no. 3, pp. 224-231, 2003.
[25] N Katoh, T. Ibaraki, and H. Mine, "An Efficient Algorithm for k Shortest Simple Paths," *Networks,* vol. 12, pp. 411-427, 1982.
[26] Y. Keselman and S. Dickinson, "Bridging the Representation Gap between Models and Exemplars," *Proc. IEEE Workshop Models versus Exemplars in Computer Vision,* Dec. 2001.
[27] Y. Keselman and S. Dickinson, "Generic Model Abstraction from Examples," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 856-863, Dec. 2001.
[28] B.B. Kimia, A. Tannenbaum, and S.W. Zucker, "Shape, Shocks, and Deformations I: The Components of Two-Dimensional Shape and the Reaction-Diffusion Space," *Int'l J. Computer Vision,* vol. 15, pp. 189-224, 1995.
[29] W.G. Kropatsch, "Building Irregular Pyramids by Dual Graph Contraction," *Proc. IEEE Vision, Image and Signal Processing,* vol. 142, no. 6, pp. 366-374, 1995.
[30] B. Leibe and B. Schiele, "Analyzing Appearance and Contour Based Methods for Object Categorization," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2003.
[31] A. Leonardis and H. Bischof, "Dealing with Occlusions in the Eigenspace Approach," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 453-458, June 1996.
[32] D. Lowe, *Perceptual Organization and Visual Recognition.* Kluwer Academic, 1985.
[33] D. Lowe, "Object Recognition from Local Scale-Invariant Features," *Proc. Int'l Conf. Computer Vision,* pp. 1150-1157, 1999.
[34] B. Luo, R. Wilson, and E. Hancock, "Spectral Embedding of Graphs," *Pattern Recognition,* 2004.
[35] A. Mohan, C. Papageorgiou, and T. Poggio, "Example-Based Object Detection in Images by Components," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 4, pp. 349-361, Apr. 2001.
[36] H. Murase and S. Nayar, "Visual Learning and Recognition of 3-D Objects from Appearance," *Int'l J. Computer Vision,* vol. 14, pp. 5-24, 1995.
[37] R. Nelson and A. Selinger, "A Cubist Approach to Object Recognition," *Proc. IEEE Int'l Conf. Computer Vision,* Jan. 1998.
[38] H. Nishida and S. Mori, "An Algebraic Approach to Automatic Construction of Structural Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 15, no. 12, pp. 1298-1311, Dec. 1993.
[39] M. Pelillo, K. Siddiqi, and S. Zucker, "Matching Hierarchical Structures Using Association Graphs," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 11, pp. 1105-1120, Nov. 1999.
[40] A. Pentland, "Perceptual Organization and the Representation of Natural Form," *Artificial Intelligence,* vol. 28, pp. 293-331, 1986.
[41] A. Pentland, "Automatic Extraction of Deformable Part Models," *Int'l J. Computer Vision,* vol. 4, pp. 107-126, 1990.
[42] A. Pope and D. Lowe, "Learning Object Recognition Models from Images," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 296-301, 1993.
[43] E. Rivlin, S. Dickinson, and A. Rosenfeld, "Recognition by Functional Parts," *Computer Vision and Image Understanding,* vol. 62, no. 2, pp. 164-176, 1995.
[44] C. Schmid and R. Mohr, "Combining Greyvalue Invariants with Local Constraints for Object Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 872-877, June 1996.
[45] S. Sclaroff and A. Pentland, "Modal Matching for Correspondence and Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, no. 6, pp. 545-561, June 1995.
[46] T.B. Sebastian, P.N. Klein, and B.B. Kimia, "Recognition of Shapes by Editing their Shock Graphs," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 5, pp. 550-571, May 2004.
[47] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* June 1997.
[48] K. Siddiqi, A. Shokoufandeh, S. Dickinson, and S. Zucker, "Shock Graphs and Shape Matching," *Int'l J. Computer Vision,* vol. 30, pp. 1-24, 1999.
[49] F. Solina and R. Bajcsy, "Recovery of Parametric Models from Range Images: The Case for Superquadrics with Global Deformations," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 12, no. 2, pp. 131-146, Feb. 1990.
[50] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience,* vol. 3, no. 1, pp. 71-86, 1991.
[51] J. Utans, "Learning in Compositional Hierarchies: Inducing the Structure of Objects from Data," *Advances in Neural Information Processing Systems,* J.D. Cowan, G. Tesauro, and J. Alspector, eds., vol. 6, pp. 285-292, Morgan Kaufmann, 1994.
[52] S. Wachsmuth, S. Stevenson, and S. Dickinson, "Towards a Framework for Learning Structured Shape Models from Text-Annotated Images," *Proc. HTL-NAACL03 Workshop Learning Word Meaning from Non-Lingustic Data,* May 2003.
[53] M. Weber, M. Welling, and P. Perona, "Unsupervised Learning of Models for Recognition," *Proc. Sixth European Conf. Computer Vision,* vol. 1, pp. 18-32, 2000.
[54] P.H. Winston, "Learning Structural Descriptions from Examples," *The Psychology of Computer Vision,* chapter 5, pp. 157-209, McGraw-Hill, 1975.
[55] Y. Xu, E. Saber, and A. Tekalp, "Dynamic Learning from Multiple Examples for Semantic Object Segmentation and Search," *Computer Vision and Image Understanding,* vol. 95, pp. 334-353, 2004.
[56] S. Zhu and A.L. Yuille, "Forms: A Flexible Object Recognition and Modelling System," *Int'l J. Computer Vision,* vol. 20, no. 3, pp. 187-212, 1996.

Yakov Keselman received the BS degree in mathematics and computer science from the Urals State University in 1991, the MS degree in mathematics from the University of Georgia in 1994, and the PhD degree in computer science from Rutgers University in 2005. He was a recepient of Rutgers University internal fellowships in digital libraries (1998–1999) and and in cognitive science (2000–2001). Since 2001, he has been an instructor at the School of Computer Science, Telecommunications, and Information Systems, DePaul University. To increase his exposure to various aspects of computing, he has taught courses in computer science, software engineering, information systems, and electronic commerce. His research interests include content-based image retrieval, applied discrete optimization, and data-driven decision making. He is a member of the IEEE and the IEEE Computer Society.

Sven Dickinson received the BASc degree in systems design engineering from the University of Waterloo in 1983 and the MS and PhD degrees in computer science from the University of Maryland in 1988 and 1991, respectively. He is currently an associate professor of computer science at the University of Toronto. From 1995-2000, he was an assistant professor of computer science at Rutgers University, where he also held a joint appointment in the Rutgers Center for Cognitive Science (RuCCS) and membership in the Center for Discrete Mathematics and Theoretical Computer Science (DIMACS). From 1994-1995, he was a research assistant professor in the Rutgers Center for Cognitive Science and, from 1991-1994, a research associate at the Artificial Intelligence Laboratory, University of Toronto. He has held affiliations with the MIT Media Laboratory (visiting scientist, 1992-1994), the University of Toronto (visiting assistant professor, 1994-1997), and the Computer Vision Laboratory of the Center for Automation Research at the University of Maryland (assistant research scientist, 1993-1994, visiting assistant professor, 1994-1997). Prior to his academic career, he worked in the computer vision industry, designing image processing systems for Grinnell Systems Inc., San Jose, California, 1983-1984, and optical character recognition systems for DEST, Inc., Milpitas, California, 1984-1985. His major field of interest is computer vision, with an emphasis on shape representation, object recognition, and mobile robot navigation. Dr. Dickinson was cochair of the 1997, 1999, and 2004 IEEE Workshops on Generic Object Recognition, while, in 1999, he cochaired the DIMACS Workshop on Graph Theoretic Methods in Computer Vision. In 1996, he received the US National Science Foundation CAREER award for his work in generic object recognition and, in 2002, he received the Government of Ontario Premiere's Research Excellence Award (PREA), also for his work in generic object recognition. From 1998-2002, he served as an associate editor of the *IEEE Transactions on Pattern Analysis and Machine Intelligence*, in which he also coedited a special issue on graph algorithms and computer vision which appeared in 2001. He currently serves as associate editor for the journal, *Pattern Recognition Letters*. He is a member of the IEEE and the IEEE Computer Society.