Deformable Shape Detection and Description via Model-Based Region Grouping

Stan Sclaroff, *Member*, *IEEE*, and Lifeng Liu

Abstract—A method for deformable shape detection and recognition is described. Deformable shape templates are used to partition the image into a globally consistent interpretation, determined in part by the minimum description length principle. Statistical shape models enforce the prior probabilities on global, parametric deformations for each object class. Once trained, the system autonomously segments deformed shapes from the background, while not merging them with adjacent objects or shadows. The formulation can be used to group image regions obtained via any region segmentation algorithm, e.g., texture, color, or motion. The recovered shape models can be used directly in object recognition. Experiments with color imagery are reported.

Index Terms—Image segmentation, region merging, object detection and recognition, deformable templates, nonrigid shape models, statistical shape models.

1 Introduction

Segmentation using a traditional low-level image processing technique, such as region growing [43], [60], requires a considerable amount of interactive guidance in order to get satisfactory results. Automating such a model-free approach is difficult because of shape complexity, illumination, interreflection, shadows, and variability within and across individual objects. In addition, noise and other image artifacts can cause incorrect regions or boundary discontinuities in objects recovered with such methods.

One solution strategy is to exploit prior knowledge to sufficiently constrain the segmentation problem. When available, such information can be used to eliminate ambiguities and reduce computational complexity in finding optimal groupings of image regions. For instance, a model-based segmentation scheme can be used in concert with image preprocessing to guide and constrain region grouping [18], [36], [41], [53]. However, the use of models in segmentation is not a panacea. Due to shape deformation and variation within object classes, a simple rigid model-based approach will break down in general. This realization has led to the use of deformable shape models in image segmentation [8], [10], [27], [29], [31], [39], [46], [48].

Another strategy is to utilize image features that are somewhat invariant to illumination [7], [25], or to directly model the physics of illumination, color, shadows, and surface interreflections [21], [24], [33], [34]. Such physically-based approaches have also been shown to improve segmentation accuracy and can be used to improve performance of model-based methods.

• *The authors are with the Image and Video Computing Group, Computer Science Department, Boston University,111 Cummington Street, Boston, MA 02215. E-mail: {sclaroff, liulf}@cs.bu.edu.*

Manuscript received 28 Apr. 1999; revised 13 Sept. 2000; accepted 24 Sept. 2000.
Recommended for acceptance by J.R. Beveridge.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 109721.

Unfortunately, the above mentioned techniques are going to make mistakes in merging regions, even in constrained contexts. This is because local constraints are in general insufficient; to gain a more reliable segmentation, global consistency must be enforced. This idea is embodied in the *principle of global coherence* [50]: The best partitioning is the one that globally and consistently explains the greatest portion of the sensed data [40], [50], [56]. The idea is consonant with the minimum description length (MDL) principle: The simplest region segmentation explaining the observations is the best [13], [30], [35], [59].

In this paper, an approach is proposed that includes two stages: oversegmentation using a traditional region segmentation algorithm, followed by deformable model-based evaluation of various region grouping hypotheses. During the second stage, region merging, deformable model fitting, and consistency checking are executed simultaneously.

A statistical shape model is used to enforce the prior probabilities on global, parametric deformations for each object class. The likelihood of a region grouping is evaluated using a cost measure that includes region compatibility, region/model area overlap, and a deformation likelihood term. The approach is general, in that it can be used to group image regions based on texture measures, color, or other image features.

Once trained, the system autonomously segments objects from the background, while not merging them with adjacent objects of similar image color. The resulting recovered parametric model descriptions can then be used directly in object recognition.

Ideally, the system should determine the optimal model-based partitioning of the image, based on the likelihood of region groupings. Unfortunately, finding *the* optimal partitioning is an NP-complete problem; therefore, approximation strategies are needed to achieve a practical system. Two different approximation strategies are evaluated for performance and accuracy: the *best first* strategy and the *global consistency* strategy. The experimental evaluation will

show that there exists a trade-off between speed and correctness.

The system has been implemented and tested in a color image segmentation application that uses 2D shape models and global deformations. In the implementation, the prior distribution on global deformations for each shape is assumed Gaussian and estimated using region segmentations provided in a training set. The system was tested on hundreds of cluttered images of objects taken from a number of different shape classes (e.g., fish, leaves, fruit, and vegetables), and results are encouraging.

2 Related Work

2.1 Segmentation via Deformable Models

Previous work in this area is based on the deformable model paradigm of active contours or *snakes* [10], [14], [28], [31], [39], [48]. Snakes incorporate prior knowledge about a contour's smoothness and resistance to deformation. A regularized estimate of a contour is obtained by defining image edge forces that "pull" on the snake model. An "internal inflation" force can be used to expand a snake past spurious edges toward real edges of the structure, making the model less sensitive to initial conditions [10].

The snake formulation can be extended to include a term that enforces homogeneous properties over the region during region growing [8], [27], [29], [46]. This hybrid approach offers the advantages of both region-based and deformable modeling techniques and tends to be more robust with respect to model initialization and noisy data. However, it requires hand-placement of the initial model or a user-specified seed point on the interior of the region. One proposed solution is to scatter many region seeds at random over the image, followed with Bayes/MDL segmentation and merging [59].

Other approaches use special-purpose deformable templates [28], [37], [55], [58]. For instance, Yuille et al. [58] employ deformable templates to model facial features, such as eyes. The template-based approach allows for inclusion of object-specific knowledge in the model. This further constrains segmentation, resulting in enhanced robustness to occlusion and noise. Furthermore, the recovered template parameters can be used for shape recognition. These methods require the careful construction and parameterization of templates.

Deformable templates can be derived semiautomatically, via statistical analysis of shape training data [12], [38]. The estimated probability density function (PDF) for the shape deformation parameters can be used in Bayesian segmentation methods.

2.2 Segmentation as Labeling

From another view, image segmentation is a labeling problem; the ideal segmentation should be consistent or nearest to the one with maximum-likelihood. This has led to various relaxation labeling or stochastic labeling methods that are related to general optimization algorithms.

Bhanu and Faugeras [4] regarded the shape matching as "segment matching" problem and maximized a criterion function based on the ambiguity and inconsistency of classification. In their approach, a contour model was used to do the matching and define the criterion function.

Hummel and Zucker [26] pointed out that a large class of problems can be formulated in terms of the assignment of labels to objects. Frequently, processes are needed that reduce ambiguity and noise and select the best labeling among several possible choices. Relaxation labeling processes are just such a class of algorithms based on the use of local constraints between labels.

Faugeras and Berthod [15] proposed a definition of a class of global criteria that combine both ambiguity and consistency and a projected gradient algorithm was developed to minimize these criteria.

Freuder [17] provided a constraint network representation for a combinatorial search problem. This method is good for reducing the search space by ruling out the inconsistent subspace; however, it does not guarantee that computational complexity of searching in the remaining space is not exponential.

Nearly all of the mentioned techniques require some prior information, such as the number of labels needed and the probability distribution of these labels in the image. However, that kind of information is not always available or is not accurate for general imagery. Without such prior information about a particular image segmentation, the system must somehow determine appropriate settings automatically.

2.3 Minimum Description Length Principle

A number of authors have proposed segmentation methods that exploit the Minimum Description Length (MDL) principle [13], [30], [35], [59]. MDL has a strong fundamental grounding, being based on information-theoretic arguments: The simplest model explaining the observations is the best and it can result in an objective function with no arbitrary thresholds.

In [35], it has been shown that a descriptive language limited to a low-order polynomial description of the region boundaries yields intuitively satisfying partitions for a wide class of images. It has been shown that by choosing optimal descriptive languages for given prior probabilities, MDL strategy is equivalent to Maximum Probability (MAP) estimation. As will be seen, the global cost function employed in our system is also compatible with the MDL principle.

Kanungo, et al. [30], presented a fast segmentation algorithm based on MDL to do multiband image segmentation. Their merging scheme is similar to highest confidence first estimation. The algorithm successively merges pairs of neighboring regions provided that the mergers decrease the total code length. At each step, the pair of regions producing the greatest code length decrease is merged.

Zhu and Yuille [59] derived a segmentation method in their FORMS system. Their method minimizes a generalized Bayes/MDL criterion using the variational principle. The algorithm combines aspects of snakes and region growing, and is guaranteed to converge to a local minimum. In related work, [23] presented a method for spatiotemporal segmentation of long sequences of images based on the MDL principle and simultaneously obtained optimal spatial segmentation and motion estimation without extracting the optic-flow field.

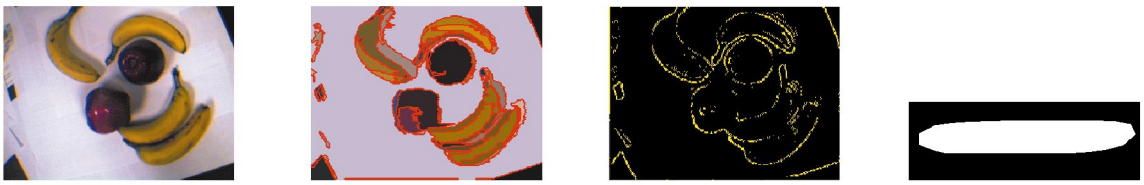


Fig. 1. Example input and precomputations: (a) input color image, (b) initial oversegmentation, (c) edge map, and (d) deformable shape template.

2.4 Optimization Algorithms

After defining the criterion function for labeling, the next problem is computing the solution to the optimization problem. Generally speaking, finding the globally-consistent image labeling is an NP-hard problem. Therefore, approximation algorithms are needed in solving any segmentation problems of realistic size. Annealing methods (simulated or deterministic) are frequently used [20].

Bongiovanni and Crescenzi [6] applied a simulated annealing method to detect ellipses and parallelograms in an image. The method seeks the combination of rigid ellipses and parallelograms that best fits the given image data. Theirs was a parallel implementation that assumed a binary image as input.

Storvik [49] used a simulated annealing scheme to detect contours via a fully Bayesian approach. Later, Grzeszczuk and Levin [22] described an image segmentation technique in which an arbitrarily shaped contour was deformed stochastically until it fitted around an object of interest. The evolution of the contour was controlled by a simulated annealing process that caused the contour to settle into the minimum of an image-derived energy function. Their method could only detect one object at a time and the correlation between objects in the image was not considered.

Wang [57] presented a segmentation method in which images are partitioned into sets of connected components. Their hierarchical approach at each step minimizes a cost function over the space of partitions on the graph of connected components. Simulated annealing is used in the minimization. This approach did not use model information and the cost function was generic; it was based on the homogeneity of the image features.

Noll and Von Seleen [42] also solved the object recognition problem by formulating an energy function that could be optimized via deterministic annealing. In addition, matching of model features was done in a coarse-to-fine manner. Compared to stochastic annealing approaches, there are some restrictions in defining an energy function for deterministic annealing, e.g., it should be easy to calculate the energy function's derivative.

Chou and Brown [9] used highest confidence first (HCF) to infer a unique labeling from the a posteriori distribution that is consistent with both the prior knowledge and evidence. Their method is analogous to deterministic annealing, but the computation is more efficient. The HCF method is closely-related to agglomerative clustering methods that were applied to image segmentation by [47].

3 The Basic Idea Behind our Approach

In our system, a deformable model is used to guide grouping of image regions. A shape model is specified in terms of global warping functions applied to a closed polygon, hereafter referred to as a template. We will now give a brief overview of the segmentation process as it is applied to find simple banana shapes in the example color image in Fig. 1a.

First, the input color image is oversegmented via standard region-merging algorithms [3], [11]. The regions found by the oversegmentation module are shown in Fig. 1b. The output of this module also includes a standard region adjacency graph. Using this oversegmentation, candidate regions for interesting objects are determined based on their color features. Using our approach, it is also possible to use texture features or intensity features to detect the candidate regions.

Next, an edge map is computed for the input image, as shown in Fig. 1c. The edge map is used to constrain consideration of possible grouping hypotheses later in region merging. Notable edges and their strengths can be detected via standard image processing methods. Alternatively, the map can be computed by segmenting the input image at various oversegmentation factors, detecting region boundaries over the various scales, and then generating a map that integrates boundary strengths over scale [11].

The system then tests various combinations of candidate region groupings. For each grouping hypothesis, we recover the model alignment and deformations needed to match the grouping. The shape template used for grouping regions in this simple example is shown in Fig. 1d. Downhill-simplex method is used to find the minimum cost configuration of the model. Our cost measure includes: 1) a region color compatibility term, 2) a region/model area overlap term, and 3) a deformation term. The deformation term enforces a priori constraints on the amounts and types of deformations allowed for a particular deformable shape class (e.g., bananas). The template is essentially an active contour that "prefers" to deform in ways that are consistent with the prior distribution on the model parameters.

Assume that there are n candidate regions for merging in an image. In theory, the system should exhaustively test all possible combinations of the n candidate regions and select the best ones for merging. In practice, region adjacency and edge map constraints are used to prune the possible region grouping hypotheses. However, the worst case computational complexity of such exhaustive testing is exponential in n and the problem of finding the best region grouping is NP-hard. To make the problem tractable, we must employ an algorithm that finds the approximately optimal region

(a) (b) (c)

Fig. 2. Example final segmentation result: (a) selected region groupings, (b) resulting model-guided region merging, and (c) recovered parametric shape models.

groupings. In our system, we have tested two general approaches that find the approximately optimal solution: best first strategy and global consistency strategy.

The "best" region grouping hypotheses selected by the best first strategy are shown in Fig. 2a. These model-guided groupings are then merged in the color image segmentation, as shown in Fig. 2b.

No matter which strategy is employed, the selected grouping hypotheses have recovered shape models associated with them, as shown in Fig. 2c. The statistical shape model allows us to estimate the likelihood that the region grouping belongs to a particular shape class. Thus, the model parameters can be used directly in recognition.

4 Deformable Model Formulation

A shape model is specified in terms of global warping functions applied to a closed polygon, hereafter referred to as a template. To deform the template, we define an N-dimensional vector of warping parameters, $\mathbf{a}$, that describe a generic deformation applied to each polygon vertex:

$$\mathbf{x}_i' = f(\mathbf{x}_i, \mathbf{a}), \tag{1}$$

where $\mathbf{x}_i$ is a vertex in the polygon before warping and $\mathbf{x}_i'$ is the vertex afterwards.

Perhaps the simplest warping functions to be used in (1) are those of a 2D affine model or an eight parameter projective model. More suitable functions for modeling general nonrigid deformation include: higher-order polynomials, orthogonal basis functions, or global deformation functions [2], [44].

One advantage of the active contours paradigm is that prior information about contour smoothness and bending can be exploited to gain a regularized estimate of the true shape. To gain a regularized solution, we minimize the strain energy incurred while deforming the model to fit the data. This results in robustness to noisy edge data and missing data.

In a traditional snakes formulation, smoothness and bending operators are defined over the *control points* of the model to obtain a stiffness matrix, $\mathbf{K}$. In a deformable template formulation, instead, we define a stiffness matrix over the *deformation parameters*. Thus, the strain energy is expressed in the template's deformation parameter space:

$$E_{strain} = \tilde{\mathbf{a}}^T \mathbf{K} \tilde{\mathbf{a}}, \tag{2}$$

where $\tilde{\mathbf{a}} = \mathbf{a} - \bar{\mathbf{a}}$ is a vector describing parameter displacement from a zero strain "rest" state (i.e., the expected value for the deformation parameters $\bar{\mathbf{a}} = E[\mathbf{a}]$).

To demonstrate the approach, we have implemented a system that uses linear and quadratic polynomials to model deformation due to stretching, shearing, bending, and tapering.

4.1 Statistical Stiffness Matrix

There is a well-understood link between physically-motivated deformable models and statistical estimation [54]. Splines were perhaps some of the first "physically-based" models employed in statistical estimation [32]; they are particularly well-suited to modeling data sampled from a Markov Random Field (MRF), with Gaussian noise added [20]. The same principles hold true for regularization [5], [54], where the energies of a physical model can be related directly with measurement and prior probabilities used in Bayesian estimation [52].

We will assume that the distribution on shape parameters for a particular shape category Ω can be adequately modeled as a multidimensional, unimodal Gaussian distribution. The distribution can be characterized by the mean $\bar{\mathbf{a}}$ and covariance matrix Σ. The likelihood of a pattern $\mathbf{a}$ is given by:

$$P(\mathbf{a}|\Omega) = \frac{\exp\left[-\frac{1}{2}\tilde{\mathbf{a}}^T \Sigma^{-1} \tilde{\mathbf{a}}\right]}{(2\pi)^{N/2}|\Sigma|^{1/2}}, \tag{3}$$

where $\tilde{\mathbf{a}} = \mathbf{a} - \bar{\mathbf{a}}$. The mean vector and covariance matrix can be obtained via a statistical analysis over a set of training shapes, as will be explained in Section 4.4.

The sufficient statistic for characterizing the likelihood is the Mahalanobis distance:

$$E_{deform} = \tilde{\mathbf{a}}^T \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{a}}, \tag{5}$$

showing the connection between the stiffness matrix $\mathbf{K}$ of (2) and the inverse covariance matrix used in the Gaussian model.

The strain energy equations of (2) and (4) can be decoupled via an eigenvector transform. In the case of the stiffness matrix formulation, this approach is known as modal analysis [44] and, in the case of the covariance matrix formulation, this is known as principal components analysis (PCA) [12]. The eigenvector transform is used to precondition the problem by diagonalizing (decoupling) the stiffness matrix. This reduces the computational complexity of evaluating (2) and (4) to be linear in the number of deformation parameters. At each step, the model parameters are recovered in the decoupled parameter space.

In our experience, the use of a Gaussian model for the prior distribution on global deformation leads to reliable shape-based image segmentation. However, if needed, the formulation of (4) can be changed to accommodate alternative statistical models, i.e.:

$$E_{deform} \propto -\log P(\mathbf{a}|\Omega), \tag{5}$$

where $P(\mathbf{a}|\Omega)$ gives the prior distribution on global deformation parameters, $\mathbf{a}$, for a particular shape class Ω.

4.2 Region Merging Cost Function

Given a list of candidate region grouping hypotheses, we need to select the most likely ones. The shape model is deformed to match each grouping hypothesis $\mathbf{g}_i$ in such a way as to minimize a cost function:

$$E(\mathbf{g}_i) = (1-\alpha)E_{color} + \alpha\big((1-\beta)E_{area} + \beta E_{deform}\big), \tag{6}$$

where α and β are scalar constants with values in the range $[0,1]$ that control the relative importance of the three terms: a region color compatibility term E_{color}, a region/model area overlap term E_{area}, and the deformation energy from (4).

The region color compatibility is related to the covariance for the pixel colors within the grouping:

$$E_{color} = \|\frac{1}{n}\sum_{j\in G}(\mathbf{c}_j - \bar{\mathbf{c}})(\mathbf{c}_j - \bar{\mathbf{c}})^T\|, \tag{7}$$

where $\mathbf{c}_j$ is a vector giving the color value at the jth pixel in the region grouping hypothesis and $\bar{\mathbf{c}}$ is the mean color vector for all regions in the candidate grouping.

The region/model area overlap term takes the form:

$$E_{area} = \frac{S_G S_m}{S_c^2}, \tag{8}$$

where S_G is the area of the region grouping hypothesis, S_m is the area of the deformed model, and S_c is the common area between the regions and deformed model. By using the degree of overlap in our cost measure, we can avoid measuring distances between region boundaries and corresponding model control points. Hence, we can avoid the problem of finding direct correspondence between landmark points, which is not easy in the presence of large deformations.

The values of the two scalar constants that control the relative importance of the three terms in (6) can be determined as follows: The value of α depends on the expected color homogeneity of the objects to be segmented in the image. Setting the parameter in the range $\alpha = [0.98, 1]$ was shown to be a good range in our experiments. In cases where the color is less important than the shape and compatability of regions, α is closer to one. The value of β depends on the confidence of the prior distribution on global deformations estimated from the training set. In our experiments, reliable segmentation was obtained with this parameter set in the range $\beta = [0.001, 0.01]$.

4.3 Model Fitting Procedure

One important step in the image partitioning procedure is to fit each region grouping hypothesis with deformable models from the object library. The fitting procedure will be used in evaluating the likelihood of each region grouping hypothesis. Therefore, it is important that the fitting procedure is efficient, fully-automatic, and reliable.

Deformable model fitting is defined in the usual way: The system must minimize a nonlinear cost function (i.e., (6)). If the gradient descent method is used to search for the optimal solution, the step size is difficult to determine reliably. In addition, the effects of the shape parameters are not independent. We have no guarantee that a traditional gradient-based minimization method will converge to the global minimum location, unless we are given an initial placement of the model that is close to the minimum already. This "initial pose problem" is a known weakness of many deformable model recovery schemes.

Approaches to solving this have been suggested: graduated nonconvexity [5], multigrid approaches [54], and nonlinear programming methods [1]. In our system, we employ the downhill-simplex method [45] because it requires only function evaluations, not derivatives. Though it is not very efficient in terms of the number of function evaluations that it requires, it is still suitable for our application since it is fully-automatic and reliable.

Downhill-simplex method must be started not just with a single point, but with $N+1$ points defining a initial simplex (where N is the number of variables in the function). For a new group of candidate regions, we can estimate the centroid, orientation, and scale factor via moments. We then sample $N+1$ rotation parameters uniformly over the range $[0, 2\pi]$ and compute the initial templates of varying orientation. The downhill-simplex algorithm is then used to adjust the parameters until convergence is obtained.

Some grouping hypotheses have considerable common regions. In such cases, the matching parameters for one grouping hypothesis can be reused for another, thereby speeding convergence.

In order to further accelerate matching, a multiple-resolution method can be employed [42]. In our case, the oversegmented image is first subsampled at various scales (without blurring). Each grouping hypothesis is first matched with the shape model in the lowest resolution image. The model fit at that resolution is then used as input at the next level of resolution, etc. In our experience, this approach significantly speeds convergence while also avoiding local minima.

4.4 Model Training

In our current system, the template is defined by the operator as a polygonal model. During model training, the system is first presented with a collection of color images. These images are first oversegmented, as described in the previous section. In the first few training images, the operator is asked to mark candidate regions that belong to the same object. The system then merges the regions and uses downhill-simplex method to minimize the cost function in (6), thereby matching the template to the training regions in a particular image. This process is repeated for all images in the training set.

In the first training images, the shape term in (6) is ignored ($\beta = 0$ in (6)). However, it is possible to extend this approach so that the system gets more independent as training progresses. As more training data is processed, the

β parameter can be increased and used to semiautomate training of the system. The system can be allowed to take a "first guess" at the correct region grouping and present it to the operator for approval. After a few training images, the system also has some idea of the "color of interest." This information is used to winnow regions considered for marking by the operator in subsequent training images.

5 Automatic Image Segmentation

Once trained, the deformable model can be used to guide grouping and merging of color regions. The process begins with oversegmentation using a traditional color region segmentation algorithm [3], [11]. There are two goals of the initial oversegmentation procedure: to avoid the effects of background and clutter in the subsequent stages and to guarantee that regions from adjacent objects are not merged. Background and clutter regions will be culled later using a model-driven approach.

Next, a region boundary map is computed for the input image. This map records the edge strength at each pixel and is used to constrain the consideration of candidate region groupings later in the segmentation process. Combining boundary information with region information also improves the accuracy and robustness of the algorithm.

Notable region boundaries and their strengths can be detected via filtering with a Laplacian of Gaussian or via steerable pyramids [16]. Alternatively, the map can be computed by segmenting the input image at various oversegmentation factors, detecting region boundaries over the various scales and, then, generating a map that integrates boundary strength over scale [11].

5.1 Candidate Region Groupings

To prime the region grouping process, candidate "interesting" regions are selected based on color characteristics, e.g., mean color, color histograms [51], normalized color measures [25], or texture features. In our current system, we use the band-rate feature [7] to detect candidate regions. The band-rate is the ratio between responses from different color channels: r/g, r/b, etc. On one hand, this feature is simple to compute and does not require illumination estimation or the gray world assumption. On the other hand, the feature still has some degree of robustness to illumination variation and color variation.

Each deformable template shape has an associated mean feature. The mean region feature vector is used to determine if the region may be part of any deformable shape models in the database (within some tolerance). This results in lists of regions that may be candidates for fitting with particular shapes.

The system then tests various combinations of candidate region groupings for each model. In theory, the system should exhaustively test all possible combinations of the candidate regions and select the best ones for merging, however, the computational complexity of such exhaustive testing is exponential and the problem of finding the best group is NP-hard. To make the problem tractable, we need to introduce further constraints on search.

In our system, there are two major constraints in the selection of candidate groupings. The first constraint is a spatial constraint: Every region in a grouping hypothesis should be adjacent to another region in the same group. The second constraint is a region boundary compatibility constraint: If the boundary between two region is "strong," then they cannot be combined in the same group.

The boundary compatibility between two regions is precomputed as follows: An edge strength accumulator s_{edge} is initialized to zero. For each of n pixels at the boundary between the two regions, the corresponding edge strengths at these pixels are added to the accumulator s_{edge}. The boundary compatibility between the two regions are then given by the equation:

$$b_{i,j} = \frac{s_{edge}}{n}. \tag{9}$$

If $b_{i,j}$ exceeds a threshold, then the pair of regions is marked as incompatible and cannot be combined into the same group. This constraint can also be embodied by deleting edges in the region adjacency graph.

Using these two constraints, we can reduce the number of grouping hypotheses that need to be tested. If need be, this number can be further reduced by considering only those groupings that include at least one region with a relatively large area.

5.2 Best First Strategy for Region Grouping

While these constraints help significantly in reducing the number of hypotheses that must be tested, still further efforts must be made to constrain and reduce the computational complexity of finding an optimal partitioning of the image. Given the complexity of the problem, it is more practical to employ algorithms that find the approximately optimal partitioning. In developing our system, a number of approximation algorithms have been implemented and compared. The simplest approach tested in our system has been the best-first strategy.

In the best-first strategy, a list of all possible grouping hypotheses is generated. Possible shape models for each hypothesis are tested based on their color band-rate feature, as described above. Once all hypotheses have been fitted with shape models, we then compare the merging cost of different grouping hypotheses, selecting the hypothesis with minimum model cost. If the cost is less than a threshold, then the regions are merged. Any hypotheses that include these merged regions are then eliminated from further consideration. If any unmerged grouping hypotheses remain, then we select the one with the minimum cost and repeat the procedure. If the cost exceeds the threshold or the hypothesis list is empty, then the procedure stops.

5.3 Alternative Strategy: Global Consistency

If the number of candidate regions in the oversegmented image is very large, the best-first strategy tends to be inefficient, e.g., it sometimes requires a number of hours to segment an image on a standard workstation (SGI R5K Indy). To obtain a practical system, an alternative strategy must be employed: global consistency.

In the global consistency strategy, for any possible partitioning of the image, we compute a global cost value for the whole configuration:

$$\mathcal{E} = (1-\gamma)\sum_{i=1}^{n} \mathbf{r}_i, \mathbf{E}(\mathbf{g}_i) + \gamma \mathbf{n}, \quad (10)$$

where γ is a constant factor, $\mathbf{n}$ is the number of the groupings in the current image partitioning, $\mathbf{r}_i$ is the ratio of $\mathbf{i}\text{th}$ group area to the total area, and $\mathbf{E}(\mathbf{g}_i)$ is the cost function for the group $\mathbf{g}_i$ (6). In our experiments, we assign $\gamma = 0.04$. Empirically, we have found that segmentation generally remains stable within the range $\gamma = [0.001, 0.4]$.

The first term in (10) is the sum of the model compatibility for every grouping in the image partition. The second term corresponds to the code length (number of models employed) and, thereby, enforces a minimum description length criterion along the lines of [35].

Searching for the globally optimal image partitioning is an NP-hard problem. As explained in the Appendix, the global cost function employed does not exhibit the optimal substructure property required for solution via dynamic programming methods. Furthermore, after the initial segmentation, the number of candidate regions is not small in general. Approximation algorithms provide a more practical solution and tend to find a near-optimal partition within a reasonable number of steps. Therefore, in our implementation, we have experimented with two common approximation algorithms in finding the global minimum cost partitioning of the image: The simulated annealing algorithm and the highest confidence first algorithm.

5.3.1 Simulated Annealing

In one implementation of our algorithm, the simulated annealing approach was employed in gaining an approximate solution [6], [19], [22], [49], [57]. In simulated annealing, the choice of the temperature sequence involves a trade-off between efficiency and convergence properties. On the one hand, choosing a sequence that decreases quickly tends to result in convergence to a local optimum. On the other hand, choosing a sequence that decreases too slowly will make the algorithm inefficient. In general, the temperature must be lowered at a very slow (logarithmic) rate in order to maintain thermal equilibrium. However, it has been shown that the temperature can be lowered more rapidly (e.g., exponentially) if moves are selected from a size distribution proportional to the Cauchy distribution [22].

In our implementation, for simplicity and effectiveness, we define the temperature to be a linear function. The temperature is initially $T = 0.4$, and after each iteration, the new temperature value is $T = T * 0.5$. The number of temperature steps in the annealing schedule is limited to twenty, and therefore $t_{min} = 0.4/(2^{20})$.

In our experience, the convergence of the simulated annealing algorithm is slow. There is an inherent trade-off between the convergence speed and the correctness of the result. This experience led us to test another approach.

5.3.2 Highest Confidence First

A deterministic algorithm, highest confidence first (HCF), can be used to improve convergence speed [9], [30]. The HCF algorithm as applied to our problem is given in Fig. 3.

In the HCF algorithm, the cost function keeps decreasing until convergence. The computational complexity is generally less than that needed to obtain similar quality segmentation results via the simulated annealing algorithm. In each HCF iteration, the number of different merging configurations tested is about $\mathbf{O}(\mathbf{n})$, where $\mathbf{n}$ is the number of regions in the image. This is because some results from the previous iteration can be reused in the next. Specifically, at each iteration (except the first), the algorithm need only compute the pairwise merging cost between all groups $\mathbf{g}_i$ and the newly-merged group from the previous iteration. As a result, the total complexity for HCF is $\mathbf{O}(\mathbf{n}^2)$.

It is possible that in some iterations, the global cost will increase; however, it will generally decrease in subsequent iterations. At each iteration, we keep the current configuration and the best configuration up until now. At termination, we can give the best configuration found during the whole procedure.

6 Examples

Each of the aforementioned optimization strategies was tested on hundreds of images from a number of different classes of cluttered color imagery: images of fruit, vegetables, and leaves collected under controlled lab conditions, and images of fish obtained from the World Wide Web. In this section, a few examples of the segmentation system performance will be shown.

The system was implemented on an SGI Indy R5K workstation. All performance statistics are reported for unoptimized code. To demonstrate the approach, a system was implemented that uses linear and quadratic polynomials to model stretching, shearing, bending, and tapering. A template for each shape class was trained, as described in Section 4.4, using between 40 and 50 examples per shape class.

The first example shows segmentation results for detecting and merging regions associated with bananas. A simple banana shape model (Fig. 1d) was trained using 40 example images of bananas at varying orientations and scales. These training images were not contained in the test image data set.

All images in the test data set were then segmented using the trained model, as described in Section 5. The best first strategy was employed in finding the best image partition.

Some example images from the test data set are shown in Fig. 4. The resulting model-based region groupings are shown below each of the original images in the figure. In cases where there were multiple yellow objects in the image, the system recovered multiple model-based groupings. Segmentation took between 30 seconds and three minutes per image.

As can be seen, the resulting segmentation in these examples is satisfactory. The system correctly grouped regions despite shadows, variation in illuminant, and shape deformation. Due to the use of model-based region merging, the system is able to avoid merging similarly colored, adjacent but separate objects. Furthermore, the approach was adept at avoiding merging objects with their similarly-colored shadows.

As explained in Section 4, each region grouping has an associated vector of shape deformation parameters $\mathbf{a}$. This vector provides a low-dimensional description of each

1. Over-segment the image using a traditional color, motion, or texture region segmentation algorithm. This yields a set of regions and a region adjacency graph.
2. Compute an edge map for the input image as described in Sec. 5.
3. For each pair of adjacent regions, compute their boundary compatibility $b_{i,j}$ via Eq. 9. If $b_{i,j}$ exceeds a threshold, then delete the corresponding edge in the adjacency graph.
4. Initialize the region grouping configuration such that every region in the over-segmented image is in its own distinct group $\mathbf{g}_i$. Save this configuration as the best found so far, C_0.
5. For each region group $\mathbf{g}_i$, fit a shape model as described in Sec. 4.3. Then compute the global cost $\mathcal{E}_o$ via Eq. 10.
6. Set $\mathcal{E}_m$ to a very large value.
7. For each pair of adjacent groups $\mathbf{g}_i, \mathbf{g}_j$ in the current configuration
   (a) Fit a single model to the combined group $\mathbf{g}_i, \mathbf{g}_j$.
   (b) Compute the global cost, $\mathcal{E}_2$ that would result if $\mathbf{g}_i, \mathbf{g}_j$ were merged.
   (c) If $\mathcal{E}_2 < \mathcal{E}_m$, then set $\mathcal{E}_m = \mathcal{E}_2$ and save this merged configuration C_m.
8. Use the merged configuration C_m as the new configuration. If $\mathcal{E}_m < \mathcal{E}_o$, then set $\mathcal{E}_o = \mathcal{E}_m$ and save this new configuration as best found so far $C_o = C_m$.
9. If C_m contains two or more adjacent region groups, then go to 6.
10. Output the best region grouping configuration C_o and the shape model for each grouping in the configuration.

Fig. 3. Model-based region grouping using the highest confidence first (HCF) approach.

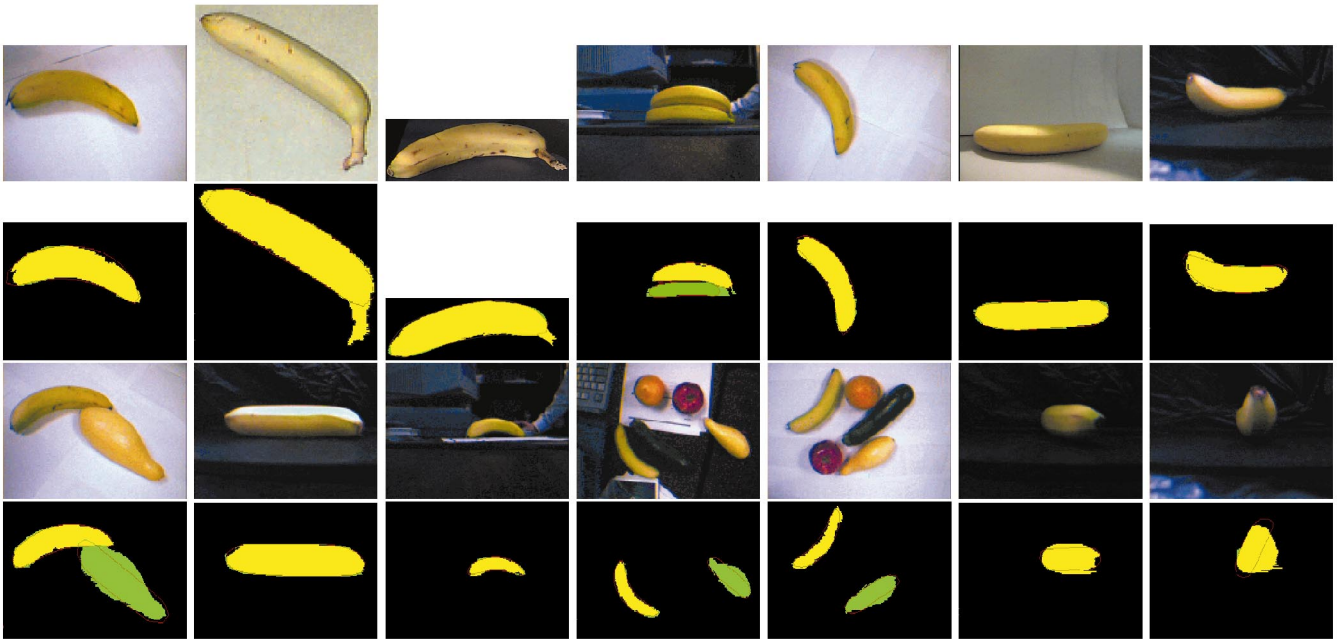


Fig. 4. Image segmentation example: Color images of bananas in various positions with varying illumination. The resulting model-based region groupings are shown below each color input image. If an image contained more than one detected shape, the shape that the system recognized as most “banana like” is labeled in yellow.

(a) (b)

Fig. 5. Two deformable template models employed in our experiments: (a) leaf model and (b) fish model. The initial polygonal model was defined by the user, and then trained as described in Section 4.4.

shape that can be stored and used for recognition. In cases where multiple objects are present, the system stores a list of model descriptions for that image.

We have conducted preliminary experiments in using these recovered shape parameter vectors for object recognition. An example is shown in Fig. 4. The user selected the image shown in the far left column. The subsequent images are shown in similarity ranking, left to right. Similarity was determined using the Mahalonobis distance between recovered $\mathbf{a}$ vectors. If there are multiple yellow objects in the input image, then the most similar shape is shown highlighted as yellow in the labeled image below. The most similar shapes are other bent bananas of similar aspect ratio. Yellow squash shapes were ranked less similar.

The next example makes use of the global consistency strategy to obtain segmentation of tropical leaf images. This example can be characterized by clutter of many simple leaves. The leaf model employed in this example was approximately an oval, as is shown in Fig. 5a. It was defined and trained as in the previous example. The training images were not contained in our test image data set. The HCF algorithm was used in finding the "best" global configuration, as described in Section 5.3.2.

Fig. 6 depicts an example of different steps in the segmentation progress using the HCF method. The original leaf image and oversegmented input are shown in Figs. 6a and 6b, respectively. Some configurations found in iterations toward the solution are shown in Figs. 6c, 6d, 6e, 6f, and 6g. In the figure, the models (shown in red) are drawn over region groupings (shown in green), therefore, the resulting areas of overlap between models and objects are shown in yellow. Fig. 6h shows the best solution found and Fig. 6i the segmentation result after merging.

The method was tested on a collection of over 100 images of different tropical leaves. Due to space limitations, not all results can be shown here. However, we include five more examples in Fig. 7. The original images are shown in the top row of the figure. The oversegmented image used as input to region merging algorithm is shown below each image (second row, Fig. 7b). The recovered shape models associated with the "best" configuration for each image are shown in the third row, Fig. 7c. Finally, the resulting model-based region merging is shown in the bottom row, Fig. 7d.

The final example shows results in segmenting images of fish obtained from the World Wide Web. The fish model shown in Fig. 5b was trained as in previous examples, using about 60 training images. The test images were excluded from the training set.

Fig. 8 shows five examples of the segmentation result obtained via the HCF optimization strategy. In nearly every case, the method accurately recovered a deformable model description of each fish in the image. Only in one case, (Fig. 8a), was the orientation of a model incorrectly estimated. Despite clutter, deformation, and partial occlusions, nearly all of the fish were accurately segmented.

7 Discussion

As seen in the examples of the previous section, the model-based region algorithm can produce satisfactory results. Based on the statistical shape model, our segmentation algorithm can detect the whole object correctly, while at the same time, avoid merging objects with background and shadow or merging adjacent multiple objects.

The major issue is computation time required to obtain a segmentation result. This led to the evaluation of different methods for obtaining approximately "optimal" region groupings. In general, each of the optimization methods offers benefits and drawbacks.

The best first strategy is a greedy algorithm. Each iteration reduces the search space and thereby avoids the combinatoric complexity in finding a globally optimal region merging. On the other hand, if the number of candidate regions in the oversegmented image is very large, the computation in the best-first strategy is still too time-consuming. In this case, the global consistency strategy can reduce search space by ruling out inconsistent configurations.

In obtaining global consistency, we implemented and compared performance for both the simulated annealing and the HCF methods. In general, simulated annealing has slower convergence speed than the HCF method. This is because in HCF, computing the merge cost value is only $\mathbf{O}(\mathbf{n}^2)$, where $\mathbf{n}$ is the total number of regions in the initially oversegmented input image.

Setting parameters in the simulated annealing approach can sometimes be problematic. The output of the simulated annealing algorithm is controlled by the annealing schedule, parameter settings, and initial grouping configuration. As described in Section 5.3.1, there is an inherent trade-off between the convergence speed (annealing schedule) and the correctness of the result.

In general, HCF converges faster than simulated annealing. However, the HCF method is not without drawbacks. First, the cost function only decreases in most cases (in general, the group number always decreases along with the iteration). The result is a locally optimal solution and not always globally optimal. Second, since HCF utilizes the best merging in every step, if the best merging during one iteration is not consistent with that in the optimal partition, then it may be lead to a wrong direction in solution space.

An example is presented in Fig. 9, where Fig. 9a is the original image and Fig. 9b is the oversegmented image, Fig. 9c is the result using simulated annealing, and Fig. 9d is the result using HCF method. Figs. 9e and 9f show the recovered models for simulated annealing and HCF,

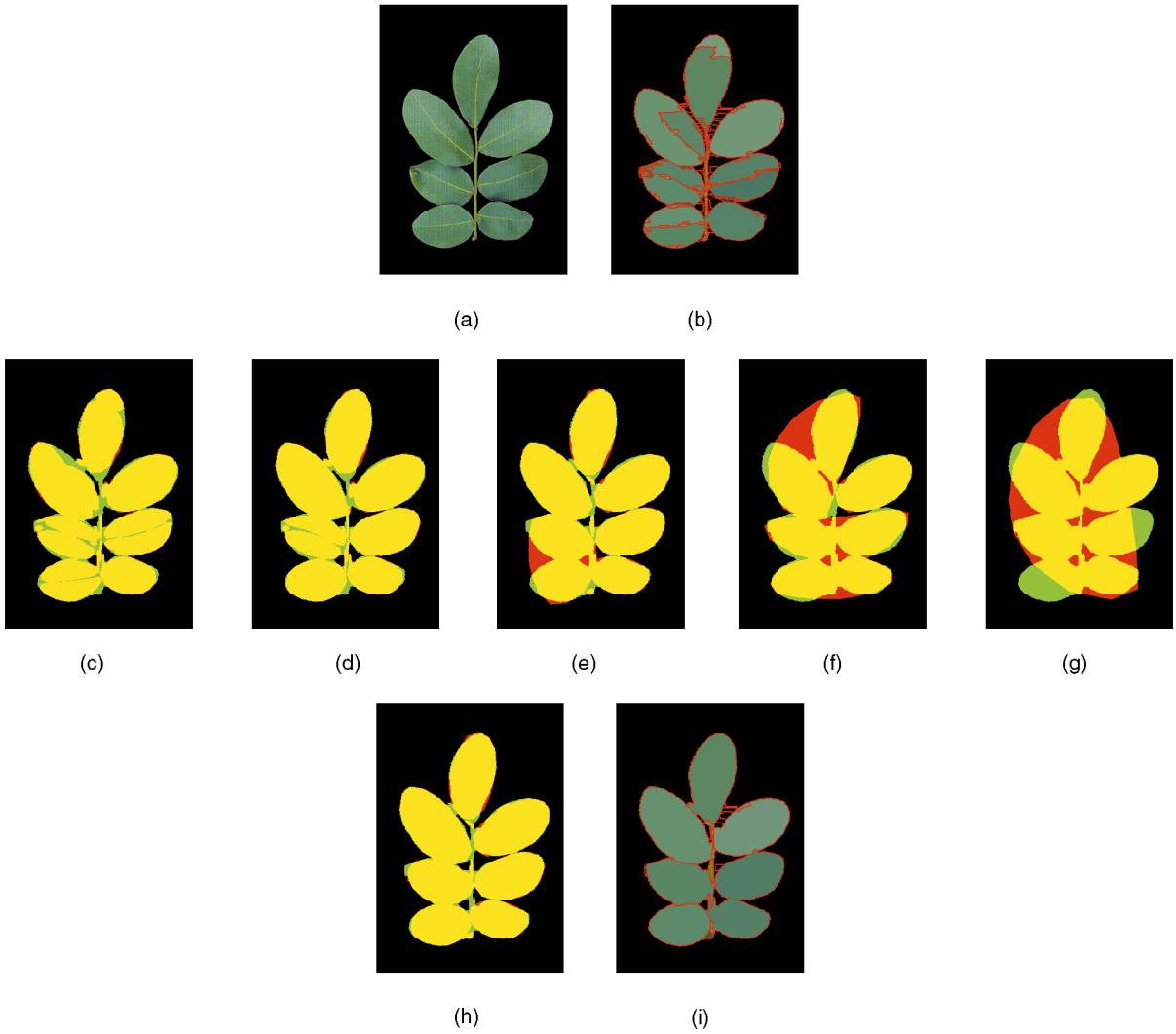


Fig. 6. Example segmentation progress using HCF method: (a) Original leaf image, (b) oversegmented image, (c), (d), (e), (f), (g) some configurations found in the evolution of the solution, (h) the best configuration found, and (i) segmentation result after merging. In (c)-(h), the models (shown in red) are drawn over region groupings (shown in green), therefore, the resulting areas of overlap between models and objects are shown in yellow.

respectively. This shows a case in which HCF could not get the best result (several leaves remain split in Fig. 9f). While simulated annealing produces the better result, it was slower to converge.

If there are shadows or partially overlapping objects in the image, then best-first strategy can get a better result since it can select the most confident group to merge first and avoid fitting spurious objects. Unfortunately, the computational complexity of best-first strategy prohibits application in general imagery.

In the simulated annealing method, the complexity is in general less than in the best first strategy. However, the degree of reduction in complexity depends on the annealing schedule and there is a trade-off between the robustness and the speed.

Therefore, the global consistency strategy (via HCF) offers a reasonable compromise between speed and accuracy. It is therefore the preferred method. As can be seen in the example segmentations of Section 6, the HCF method is able to obtain a satisfactory segmentation despite clutter, variation in illuminant, shape deformation, etc.

8 Conclusion

In previous approaches to deformable model-based segmentation, initial model placement is either given by the operator, or by exhaustively testing the model in all orientations, scales, and deformations centered at every pixel in the image. The region-based approach proposed in this paper significantly reduces the need to test all model positions.

The method includes two stages: Oversegmentation using a traditional region segmentation algorithm, followed by deformable model-based region merging via grouping and model hypothesis selection. During the second stage, region merging and object identification are executed simultaneously. A statistical shape model is used to enforce the prior probabilities on global, parametric deformations for each object class. Once trained,

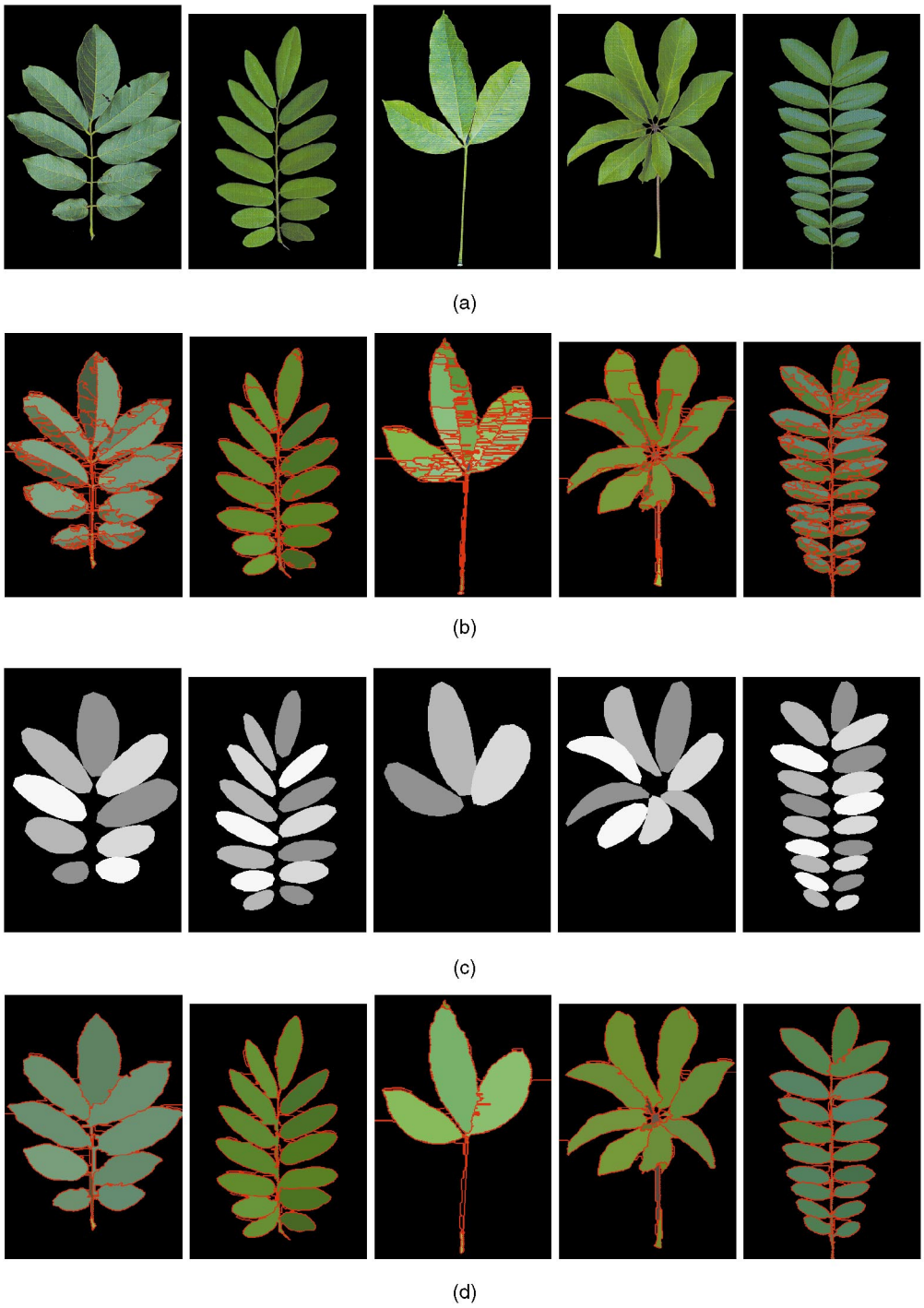


Fig. 7. Example segmentation for leaf images: (a) Original leaf image and (b) oversegmented image. The shape models recovered for the best region merging configuration are shown in (c). The model-based region merging result is shown in (d).

the system autonomously segments deformed shapes from the background, while not merging them with adjacent objects or shadows. The formulation is general, in that it can be used to group image regions obtained via any region segmentation algorithm, e.g., texture, color, or motion.

There are two kinds of strategies that were tested in the merging stage. The *best-first strategy* is suitable for images with major occlusion or partially overlap, however, its computation complexity requires that the number of candidate regions after the oversegmentation is not large. This makes the best-first strategy impractical for general application.

An alternative strategy, *global consistency* must therefore be employed. In this approach, a global cost function is employed in finding the globally-consistent region merging for the image. The global cost function is formulated so as to exploit the minimum description length (MDL) principle. Finding a globally-consistent segmentation requires the use of optimization algorithms. Two global

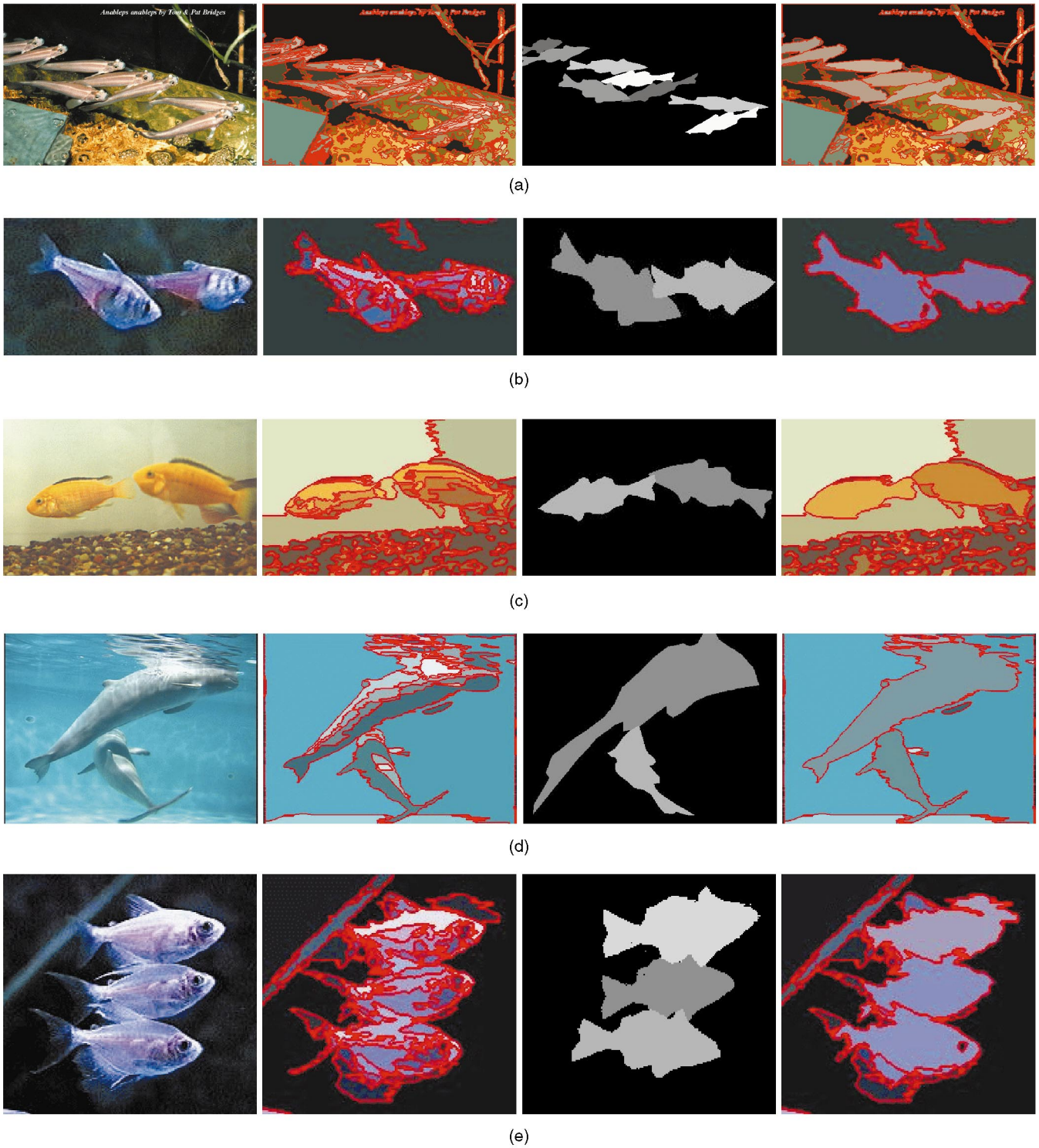


Fig. 8. Example segmentation for images of fish. The original images are shown in the first column, followed by the oversegmented images used as input to the merging algorithm. The third column shows the models selected in the best merging configuration obtained via HCF. Finally, the last column depicts the model-based merging.

optimization strategies were evaluated: simulated annealing and highest confidence first (HCF). In our experience so far, the HCF method is fast and can get a good result in most cases. However, as suggested by other authors [6], [13], [15], the time needed to compute a solution to any of the above optimization problems can be further improved via parallel algorithms.

Perhaps the major limitation of our current method is that it cannot handle large occlusions. Our next goal is to incorporate a mixture model in our system to model overlapping objects. Issues of computational complexity were addressed through the use of constraints as was described in Section 5 and the use of multiscale segmentation. However,

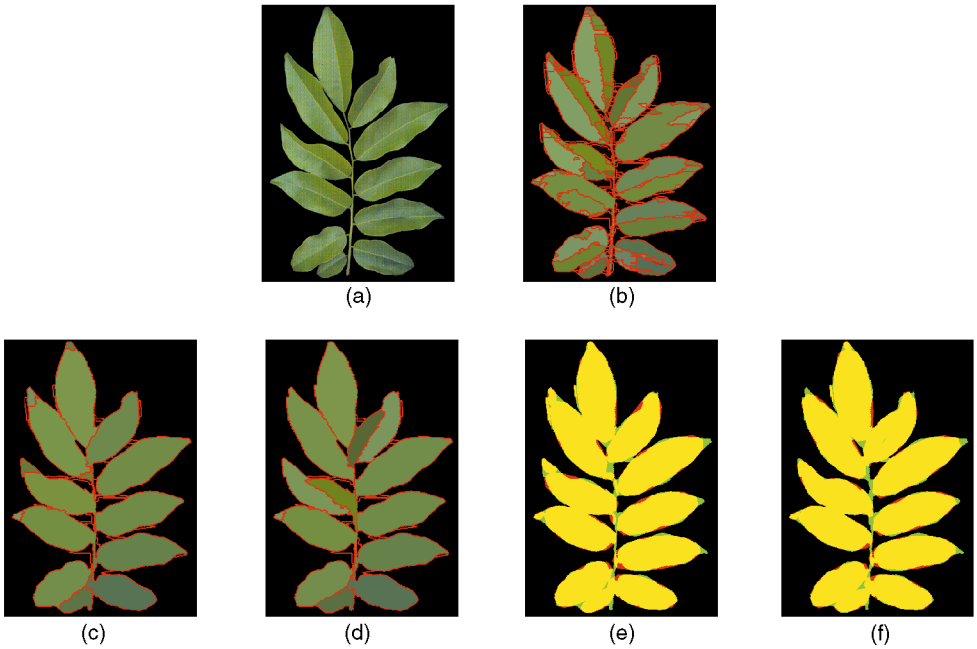


Fig. 9. Comparison segmentation results for leaf image using simulated annealing and HCF methods: (a) Original leaf image, (b) oversegmented image, (c) merge result by simulated annealing method, (d) merge result using HCF method, (e) models recovered by simulated annealing, (f) models recovered by HCF. In (e) and (f), models (shown in red) are drawn over region groupings (shown in green), therefore, the resulting areas of overlap between models and objects are shown in yellow.

the worst-case complexity is still daunting in cluttered imagery and needs to be improved.

Experiments in color image segmentation were reported and the results are encouraging. Based on the statistical shape model, our segmentation algorithm can detect the whole object correctly, while at the same time, avoid merging objects with background and shadow or merging adjacent multiple objects. Each selected grouping hypothesis has a recovered shape model associated with it, thus, the model parameters can be used directly in recognition.

APPENDIX

GLOBAL COST FUNCTION PROPERTIES

The optimal region grouping problem can be understood if we represent the image with a planar graph $\mathbf{G}$. Every image region is represented by a vertex in the graph. There is an edge between the ith vertex and jth vertex if the ith region is adjacent to the jth region in the image. In segmentation, the goal is to partition the graph so that every group is a connected subgraph, and such that the partition minimizes the global cost value of (10).

The global cost function includes two terms: The first term is the normalized sum of $\mathbf{E}(\mathbf{g}_i)$, i.e., sum of

$$\mathbf{E}(\mathbf{g}_i)\frac{\mathbf{area}(\mathbf{g}_i)}{\sum_{j=1}^{n}\mathbf{area}(\mathbf{g}_j)},$$

where $\mathbf{n}$ is the number of groups in the current configuration. Normalization can reduce the effect of small groups on the final solution. It gives priority to improving the quality of larger region groups during the evolution, and makes the system more robust since it tends to avoid the influence of noise.

The second term in (10), increases with the number of groups in the partition. This term is used to make the solution obey the minimum description length (MDL) principle. Because of the second term, the cost function does not satisfy the optimal substructure property required of dynamic programming methods. A problem is said to exhibit optimal substructure if an optimal solution to the problem contains within it optimal solutions to subproblems.

The proof is shown as follows: Assume $\mathbf{O}$ is the optimal parition for the whole image $\mathbf{I}$. If we write $\mathbf{O} = \mathbf{O}_1 + \mathbf{O}_2$ (disjoint union), where $\mathbf{O}_1$ and $\mathbf{O}_2$ are solutions corresponding to subimage $\mathbf{I}_1$ and $\mathbf{I}_2$, respectively ($\mathbf{I}$ is the disjoint union of $\mathbf{I}_1$ and $\mathbf{I}_2$). However, $\mathbf{O}_1$ may not be the optimal partitioning for the subimage $\mathbf{I}_1$. The cost value for $\mathbf{O}_1$ in subimage $\mathbf{I}_1$ is:

$$\mathcal{E}_1 = \sum_{i=1}^{n_1}\left(\frac{\mathbf{area}(\mathbf{g}_i)}{\mathbf{area}(\mathbf{I}_1)}\mathbf{E}(\mathbf{g}_i)\right) + \gamma\mathbf{n}_1, \tag{11}$$

where $\mathbf{n}_1$ is the number of groups in $\mathbf{O}_1$.

Similarly, the cost value for $\mathbf{O}_2$ in subimage $\mathbf{I}_2$ is:

$$\mathcal{E}_2 = \sum_{i=1}^{n_2}\left(\frac{\mathbf{area}(\mathbf{g}_i)}{\mathbf{area}(\mathbf{I}_2)}\mathbf{E}(\mathbf{g}_i)\right) + \gamma\mathbf{n}_2, \tag{12}$$

where $\mathbf{n}_2$ is the number of groups in $\mathbf{O}_2$.

The cost value for $\mathbf{O}$ in the image $\mathbf{I}$ is:

$$\mathcal{E} = \sum_{i=1}^{n_1+n_2}\left(\frac{\mathbf{area}(\mathbf{g}_i)}{\mathbf{area}(\mathbf{I}_1)+\mathbf{area}(\mathbf{I}_2)}\mathbf{E}(\mathbf{g}_i)\right) + \gamma(\mathbf{n}_1+\mathbf{n}_2). \tag{13}$$

Therefore, $\mathcal{E}$ is not the linear combination of $\mathcal{E}_1$ and $\mathcal{E}_2$. Even if $\mathbf{O} = \mathbf{O}_1 + \mathbf{O}_2$ is the optimal solution in image $\mathbf{I}$

($\mathbf{I} = \mathbf{I}_1 + \mathbf{I}_2$), $\mathbf{O}_1$ and $\mathbf{O}_2$ may not be the optimal solution in the subimages $\mathbf{I}_1$ and $\mathbf{I}_2$.

The ideal properties for the global cost function are: 1) satisfy the optimal substructure property and 2) scale invariance. By scale invariance, we mean that, if the image is scaled, then the optimal solution doesn't change. If we change the cost function to be

$$\mathcal{E} = \sum_{i=1}^{n} (\mathbf{area}(\mathbf{g}_i)\mathbf{E}(\mathbf{g}_i)) + \gamma \mathbf{n}, \tag{14}$$

then it satisfies optimal substructure, but it is not scale invariant. For different scales, the first term changes while the second term remains fixed. Therefore, it is possible that at one scale $\mathcal{E}(\mathbf{O}) < \mathcal{E}(\mathbf{O}')$, but at another scale $\mathcal{E}(\mathbf{O}) > \mathcal{E}(\mathbf{O}')$.

REFERENCES

[1] A.A. Amini, T.E. Weymouth, and R.C. Jain, "Using Dynamic Programming for Solving Variational Problems in Vision," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 12, no. 9, pp. 855-867, Sept. 1990.

[2] A. Barr, "Global and Local Deformations of Solid Primitives," *Computer Graphics,* vol. 18, no. 3, pp. 21-30, 1984.

[3] J.R. Beveridge, J.S. Griffith, R.R. Kohler, A.R. Hanson, and E.M. Riseman, "Segmenting Images Using Localized Histograms and Region Merging," *Int'l J. Computer Vision,* vol. 2, no. 3, pp. 311-352, Jan. 1989.

[4] B. Bhanu and O.D. Faugeras, "Shape Matching of Two-Dimensional Objects," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 6, no. 2, pp. 137-156, Feb. 1984.

[5] A. Blake and A. Zisserman, *Visual Reconstruction.* M.I.T. Press, 1987.

[6] G. Bongiovanni and P. Crescenzi, "Parallel Simulated Annealing for Shape Detection," *Computer Vision and Image Understanding,* vol. 61, no. 1, pp. 60-69, Jan. 1995.

[7] M.H. Brill, "Can Color-Space Transformation Improve Color Constancy Other than Von Kries?" *Proc. SPIE, Human, Vision, Visual Processing, and Digital Display TV,* vol. 1913, pp. 485-492, 1993.

[8] A. Chakraborty, L.H. Staib, and J. Duncan, "Deformable Boundary Finding Influenced by Region Homogeneity," *Proc. Computer Vision and Pattern Recognition,* pp. 624-627, 1994.

[9] P.B. Chou and C.M. Brown, "The Theory and Practice of Bayesian Image Labeling," *Int'l J. Computer Vision,* vol. 4, pp. 185-210, 1990.

[10] L.D. Cohen, "On Active Contour Models and Balloons," *Proc. CVGIP: Image Understanding,* vol. 53, no. 2, pp. 211-218, Mar. 1991.

[11] D. Comaniciu and P. Meer, "Robust Analysis of Feature Spaces: Color Image Segmentation," *Proc. Computer Vision and Pattern Recognition,* pp. 750-755, 1997.

[12] T.F. Cootes, A. Hill, C.J. Taylor, and J. Haslam, "Use of Active Shape Models for Locating Structure in Medical Images," *Image and Vision Computing,* vol. 12, no. 6, pp. 355-365, July 1994.

[13] T. Darrell and A.P. Pentland, "Cooperative Robust Estimation Using Layers of Support," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, no. 5, pp. 474-487, May 1995.

[14] A. DelBimbo and P. Pala, "Visual Image Retrieval by Elastic Matching of User Sketches," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 2, pp. 121-132, Feb. 1997.

[15] O.D. Faugeras and M. Berthod, "Improving Consistency and Reducing Ambiguity in Stochastic Labeling: An Optimization Approach," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 3, no. 4, pp. 412-424, Apr. 1981.

[16] W. Freeman and E.H. Adelson, "The Design and Use of Steerable Filters," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 13, no. 9, pp. 891-906, Sept. 1991.

[17] E.C. Freuder, "Synthesizing Constraint Experssions," *Comm. ACM,* vol. 21, no. 11, pp. 958-966, 1978.

[18] P. Fua and A.J. Hanson, "Using Generic Geometric Models for Intelligent Shape Extraction," *Proc. Am. Assoc. for Artificial Intelligence Conf. '87,* pp. 706-711, 1987.

[19] D. Geman, S. Geman, C. Graffigne, and P. Dong, "Boundary Detection by Constrained Optimization," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 12, no. 7, pp. 609-628, July 1990.

[20] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distribution, and Bayesian Restoration of Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 6, no. 11, Nov. 1984.

[21] R. Gershon, A.D. Jepson, and J.K. Tsotsos, "Ambient Illumination and the Determination of Material Changes," *J. Optical Soc. of Am. A,* vol. 3, no. 10, pp. 1700-1707, 1986.

[22] R.P. Grzeszczuk and D.N. Levin, "Brownian Strings: Segmentating Images with Stochastically Deformable Contours," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 10, pp. 1100-1114, Oct. 1997.

[23] H. Gu, Y. Shirai, and M. Asada, "MDL-Based Spatiotemporal Segmentation from Motion in a Long Image Sequence," *Proc. IEEE Computer Vision and Pattern Recognition Conf.,* pp. 448-453, 1994.

[24] G. Healey, "A Color Reflectance Model and Its Use for Segmentation," *Proc. Int'l Conf. Computer Vision '88,* pp. 460-466, 1988.

[25] G. Healey, "Segmenting Images Using Normalized Color," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 22, pp. 64-73, 1992.

[26] R.A. Hummel and S.W. Zucker, "On the Foundations of Relaxation Labeling Processes," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 5, no. 3, pp. 267-287, Mar. 1983.

[27] J. Ivins and J. Porrill, "Active-Region Models for Segmenting Textures and Colors," *Image and Vision Computing,* vol. 13, no. 5, pp. 431-438, June 1995.

[28] A.K. Jain, Y. Zhong, and S. Lakshmanan, "Object Matching Using Deformable Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 3, pp. 267-278, Mar. 1996.

[29] T.N. Jones and D.N. Metaxas, "Segmentation Using Deformable Models with Affinity-Based Localization," *Proc. First Joint Conf. Computer Vision, Virtual Reality, and Robotics in Medicine and Medical Robotics and Computer-Assisted Surgery (CVRMed-MRCAS '97),* pp. 53-62, 1997.

[30] T. Kanungo, B. Dom, W. Niblack, and D. Steele, "A Fast Algorithm for MDL-Based Multi-Band Image Segmentation," *Proc. IEEE Computer Vision and Pattern Recognition Conf.,* pp. 609-616, 1994.

[31] M. Kass, A.P. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *Int'l J. Computer Vision,* vol. 1, no. 4, pp. 321-331, Jan. 1988.

[32] G. Kimeldorf and G. Wahba, "A Correspondence between Bayesian Estimation and on Stochastic Processes and Smoothing by Splines," *Annals of Math. Statistics,* vol. 41, no. 2, pp. 495-502, Feb. 1970.

[33] G.J. Klinker, S.A. Shafer, and T. Kanade, "Image Segmentation and Reflection Analysis through Color," *Int'l J. Computer Vision,* vol. 2, no. 1, pp. 7-32, June 1988.

[34] G.J. Klinker, S.A. Shafer, and T. Kanade, "A Physical Approach to Color Image Understanding," *Int'l J. Computer Vision,* vol. 4, no. 1, pp. 7-38, Jan. 1990.

[35] Y.G. Leclerc, "Constructing Simple and Stable Descriptions for Image Partitioning," *Int'l J. Computer Vision,* vol. 3, pp. 73-102, 1989.

[36] M.D. Levine and S.I. Shaaheen, "A Modular Computer Vision System for Picture Segmentation and Interpretation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 3, no. 5, pp. 540-556, Sept. 1981.

[37] K.V. Mardia, W. Qian, D. Shah, and K.M.A. Desouza, "Deformable Template Recognition of Multiple Occluded Objects," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 9, pp. 1035-1042, Sept. 1997.

[38] J. Martin, A. Pentland, S. Sclaroff, and R. Kikinis, "Characteristics of Neuropathological Shape Deformations," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 30, no. 2, pp. 91-112, Feb. 1998.

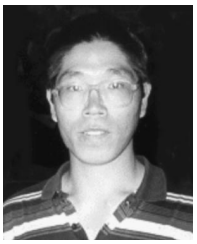
[39] T. McInerney and D. Terzopoulos, "Topologically Adaptable Snakes," *Proc. Int'l Conf. Computer Vision,* pp. 840-845, 1995.

[40] D.M. McKeown, W.A. Harvey, and J. McDermott, "Rule Based Interpretation of Aerial Imagery," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 7, no. 5, pp. 570-585, Sept. 1985.

[41] M. Nagao, T. Matsuyama, and Y. Ikeda, "Region Extraction and Shape Analysis in Aerial Photographs," *Computer Graphics and Image Processing,* vol. 10, pp. 195-223, 1979.

[42] D. Noll and W. Von Seelen, "Object Recognition by Deterministic Annealing," *Image and Vision Computing,* vol. 15, no. 11, pp. 855-860, Nov. 1997.
[43] T. Pavlidis, "Segmentation of Pictures and Maps through Functional Approximation," *Computer Graphics and Image Processing,* vol. 1 pp. 360-372, 1972.
[44] A. Pentland, "Automatic Extraction of Deformable Part Models," *Int'l J. Computer Vision,* vol. 4, no. 2, pp. 107-126, Mar. 1990.
[45] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes in C.* Cambridge, UK: Cambridge Univ. Press, 1988.
[46] R. Ronfard, "Region-Based Strategies for Active Contour Models," *Int'l J. Computer Vision,* vol. 13, no. 2, pp. 229-251, Oct. 1994.
[47] J. Silverman and D. Cooper, "Bayesian Clustering for Unsupervised Estimation of Surface and Texture Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 10, no. 4, pp. 482-495, Apr. 1988.
[48] L.H. Staib and J.S. Duncan, "Boundary Finding with Parametrically Deformable Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 14, no. 11, pp. 1061-1075, Nov. 1992.
[49] G. Storvik, "Bayesian Approach to Dynamic Contours through Stochastic Sampling and Simulated Annealing," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 16, no. 10, pp. 976-986, Oct. 1994.
[50] T.M. Strat, *Natural Object Recognition.* Springer-Verlag, 1992.
[51] M. Swain and D. Ballard, "Color Indexing," *Int'l J. Computer Vision,* vol. 7, no. 1, pp. 11-32, 1991.
[52] R. Szeliski, *Bayesian Modeling of Uncertainty in Low-Level Vision.* Kluwer Academic, 1989.
[53] J.M. Tenenbaum and H.G. Barrow, "Experiments in Interpretation Guided Segmentation," *Artificial Intelligence,* vol. 8, pp. 241-274, 1977.
[54] D. Terzopoulos, "Regularization of Inverse Visual Problems Involving Discontinuities," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 8, no. 4, pp. 413-424, July 1986.
[55] D. Terzopoulos, "On Matching Deformable Models to Images: Direct and Iterative Solutions," *Topical Meeting on Machine Vision,* vol. 12, pp. 160-167, 1987.
[56] D. Waltz, "Generating Semantic Descriptions from Drawings of Scenes with Shadows," PhD thesis, Massachusetts Inst. of Technology, 1972.
[57] J.-P. Wang, "Stochastic Relaxation on Partitions with Connected Components and Its Application to Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 6, pp. 619-636, June 1998.
[58] A.L. Yuille, D.S. Cohen, and P.W. Hallinan, "Feature Extraction from Faces Using Deformable Templates," *Int'l J. Computer Vision,* vol. 8, no. 2, pp. 99-111, Aug. 1992.
[59] S.C. Zhu and A. Yuille, "Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 9, pp. 884-900, Sept. 1996.
[60] S.W. Zucker, "Region Growing: Childhood and Adolescence," *Computer Graphics and Image Processing,* vol. 5, pp. 382-399, 1976.

Stan Sclaroff received the SM degree and the PhD degree from the Massachusetts Institute of Technology, Cambridge Massachusetts, in 1991 and 1995, respectively. He is an assistant professor in the Computer Science Department at Boston University, where he founded the Image and Video Computing Research Group. In 1996, he received a Young Investigator Award from the US Office of Naval Research and a Faculty Early Career Development Award from the US National Science Foundation. From 1989 to 1994, he was a research assistant in the Vision and Modeling Group at the MIT Media Laboratory. Prior to that, he worked as a senior software engineer in the solids modeling and computer graphics groups at Schlumberger Technologies, CAD/CAM Division. Dr. Sclaroff has coauthored more than 50 refereed publications in the areas of deformable shape modeling, image and video database retrieval, image segmentation, human tracking, and computer animation. He is currently an associate editor of the *IEEE Transactions on Pattern Analysis and Machine Intelligence*. He is a member of the IEEE.



Lifeng Liu received the BS and MS degrees in computer science from Tsinghua University, Beijing, China, in 1992 and 1996, respectively. He is a PhD candidate in the Computer Science Department at Boston University. His research interests are in computer vision, image processing, and pattern recognition, including model-based matching, image segmentation, object detection and classification, medical image database retrieval, etc.