

MosaicShape: Stochastic Region Grouping with Shape Prior

Jingbin Wang[†] Erdan Gu[‡] Margrit Betke[†]

[†]Computer Science Department, Boston University, MA, 02215

[‡]Computer and Information Science Department, University of Pennsylvania, PA, 19104

Abstract

A method that combines shape-based object recognition and image segmentation is proposed for shape retrieval from images. Given a shape prior represented in a multi-scale curvature form, the proposed method identifies the target objects in images by grouping oversegmented image regions. The problem is formulated in a unified probabilistic framework, and object segmentation and recognition are accomplished simultaneously by a stochastic Markov Chain Monte Carlo (MCMC) mechanism. Within each sampling move during the simulation process, probabilistic region grouping operations are influenced by both the image information and the shape similarity constraint. The latter constraint is measured by a partial shape matching process. A generalized cluster sampling algorithm [1], combined with a large sampling jump and other implementation improvements, greatly speeds up the overall stochastic process. The proposed method supports the segmentation and recognition of multiple occluded objects in images. Experimental results are provided for both synthetic and real images.

1 Introduction

The general goal of image segmentation is to group pixels in an image and thus produce contours or regions that are consistent with what a human observer would perceive. Segmentation methods are particularly useful if they yield object recognition. Objects with similar shapes that belong to the same object class may appear differently in images due to differing surface properties, i.e., colors and textures, or illumination conditions. Segmentation becomes particularly challenging when the object is present in cluttered environments that include other objects. A strategy that addresses these challenges is to incorporate high-level prior knowledge about the object of interest, such as a shape prior, into the segmentation process.

To capture the boundary of an object, contour-based segmentation methods, e.g., active contour or level set methods [5, 9], explicitly or implicitly deform the contour. In other work [7, 3], learned shape priors were introduced to constrain the 2D (3D) contour (surface) deformations so that objects with these predefined boundary shapes could

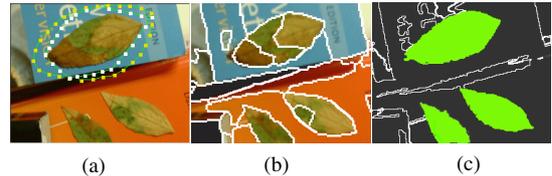


Figure 1: (a): Contour-based segmentation by the traditional active contour method [5], where the initial and final contour are shown as the link of yellow and white knots, respectively. (b): Region-based segmentation by data-driven MCMC [17]. (c): Segmentation by the proposed method.

be extracted from cluttered backgrounds. The original limitations of contour-based methods, i.e., the initialization and local minima problems (Fig. 1a), still remained.

Compared to the contour-based methods, recent region-based methods [17, 12, 6] have the following advantages. First, region-based methods are bottom up and data driven. They generally do not require an initialization step, and are expected to converge to the globally optimal solution in many cases. Second, different types of high-level prior knowledge, such as color/texture models [17, 12], boundary continuity hypotheses [17, 12], or perceptual measurements [6], can be incorporated into the bottom-up segmentation process. Application of region-based methods was mainly restricted to generate perceptual groupings or visually pleasing segmentation results (Fig. 1b). This paper proposes a novel framework that aims beyond these goals – the proposed method generates segmentation results that can be directly used for object recognition. The main contributions are:

- Given prior knowledge of shape, the proposed method identifies target objects in images by grouping oversegmented image regions via a bottom-up Markov Chain Monte Carlo (MCMC) mechanism. This method achieves simultaneous segmentation and recognition of multiple occluded objects whose location, rotation, and scale are unknown Fig. 1c).
- During the stochastic simulation process, probabilistic region grouping operations are influenced by both the image information and the shape similarity constraint.
- A great speedup of the segmentation process is gained by carefully adapting Barbu and Zhu’s [1] cluster sam-

pling method, the *Swendsen-Wang Cut* (SWC) algorithm, to the current problem and providing new implementation improvements.

The work most relevant to the current method was proposed by Sclaroff et al. [14]. The method learned a deformable shape template from multiple shape samples and applied it to constrain the region grouping process. Objects were identified by merging the regions of interest gradually, either with simulated annealing or a greedy process. Recent work by Tu et al. [16] advanced in the direction of accomplishing object segmentation and recognition simultaneously within a unified framework [17]. Another idea of combining top-down and bottom-up segmentation was demonstrated recently by Borenstein et al. [2] for segmenting the foreground objects from images. These systems [16, 2] required the feature-based prior models to be carefully constructed through a learning process, and then applied the image information as the only cues for grouping regions. These methods also did not handle occluded objects explicitly.

2 Problem Definition

Given an input color image (Fig. 2a), an “oversegmentation” (Fig. 2b) can be obtained by existing segmentation methods, e.g., the “mean shift method” [4]. Because an object of interest may be partitioned into multiple atomic regions in the oversegmented image, the oversegmentation cannot directly provide meaningful object interpretations. In our method, a shape prior is introduced (Fig. 2c), and a meaningful segmentation can be achieved where the objects of the interest in the image are identified (Fig. 2d).

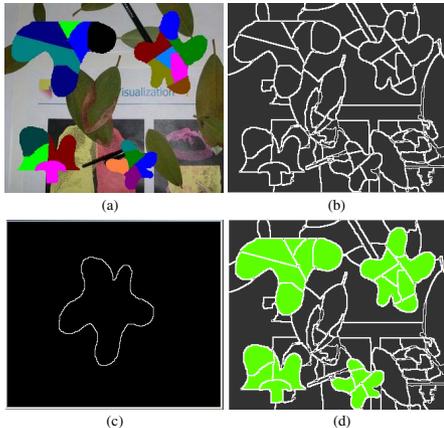


Figure 2: (a): Input color image. (b): Boundaries of oversegmented regions [4]. (c): Shape prior. (d): Recognized objects of interest, including completely or partially matched shape.

The segmentation problem to be solved here is, given a shape prior represented in some form, how to group the

atomic image regions of an oversegmented image such that the objects similar to the shape prior can be correctly identified. Finding the optimal solution for such a grouping problem is NP-hard [14]. A top-down shape matching process could be extremely slow if an exhaustive search for the object location, rotation and scale is employed. Another problem is that an image region may only provide a partial hypothesis for shape similarity. A partial region boundary that matches well locally may not produce a good match with the shape prior globally.

3 Multi-Scale Curvature Representation of Shape Prior

Defining a good representation of shape is challenging [8, 18]. For the problem at hand, the representation of the shape prior must be translation, rotation, and scale invariant. The proposed method uses boundary curvatures to define the shape of an object. Given the 2D closed contour of an object, parameterized by the arc length parameter as $\{(x(u), y(u)) | u \in [0, 1]\}$, its smoothed boundary $(x(u, \sigma), y(u, \sigma))$ can be calculated [10] by convolving the boundary curve with different sizes of Gaussian kernels, where σ is the width of the kernel. The curvature of this smoothed boundary is then:

$$\kappa(u, \sigma) = \frac{x'(u, \sigma)y''(u, \sigma) - x''(u, \sigma)y'(u, \sigma)}{(x'^2(u, \sigma) + y'^2(u, \sigma))^{3/2}}. \quad (1)$$

To achieve a scale invariant representation, Ref. [10] normalized the curvature values by the length of the contour. This approach would be problematic for handling partial shape matches as described below. The proposed method therefore instead precomputes a set of boundary curvatures for the object in different scales. In particular, a shape prior $S = \{C_{i, \sigma_j}^*\}$ consists of sequences of curvature values C_{i, σ_j}^* along the boundary of the object, smoothed by a Gaussian kernel of width σ_j , in the i th level of scale. The total number of such curvature sequences are $m \times n$, for $i = 1, \dots, m$ different levels of scale and $j = 1, \dots, n$ different sizes of Gaussian kernels.

3.1 Partial Shape Matching Problem

The boundaries of oversegmented image regions typically only partially match with the shape prior contour (Fig. 2b). To apply the prior shape information to group oversegmented image regions, partial shape similarity between the image regions and the shape prior needs to be measured. However, finding a general solution for identifying the matches of different parts of the shapes is still an unsolved problem [18, 11]. For the current problem, we assume the matched portions of the boundary of the same object are connected and only small “gaps” are allowed between them.

PARTIALSHAPEMATCH (C_V, C^*, T_κ, T_g)
// C_V : curvatures of input object; C^* : curvatures of prior shape;
// T_κ : curvature similarity threshold; T_g : gap size threshold;
 $Maxhits = 0$;
for $i = 1$ **to** $\ell(C^*)$ **do**
 $Hits = Nonhits = 0$;
 for $j = 1$ **to** $2\ell(C_V)$ **do**
 $m = j$; $n = i + j$;
 if $j > \ell(C_V)$ **then** $m = j - \ell(C_V)$;
 if $i + j \geq \ell(C^*)$ **then** $n = i + j - \ell(C^*)$;
 if $|C(m) - C^*(n)| < T_\kappa$ **then** $Hits++$;
 else $Nonhits++$;
 if $Nonhits / Hits > T_g$ **then**
 if $Hits > Maxhits$ **then** $Maxhits = Hits$;
 $Nonhits = Hits = 0$;
 $Normhits = Maxhits / \ell(C_V)$;
Return $Normhits$;

Given an input object V , for instance, an image region, and a shape prior S , their boundary curvature sequences are C_V for V and C^* for S (Eq. 1). The proposed PARTIALSHAPEMATCH algorithm (*PSM*) identifies the longest subsequence that matches in both C_V and C^* . It allows small gaps in the final matched sequences, controlled by threshold T_g . The algorithm’s nested loop repeats $2\ell(C_V)$ times to allow the matching to start from any position on the region boundary due to its cyclic representation. The length ℓ of the output subsequence is normalized by the boundary length of the input object V . The shape similarity

$$M(V, S) = \max\{d_{i,j,k} = PSM(C_{V,\sigma_i}, C_{j,\sigma_k}^*), \forall i, j, k\} \quad (2)$$

between V and S requires *PSM* to be performed for all scale and smoothness levels, where C_{V,σ_i} is the smoothed boundary curvature for image region V .

We assume the length of the region boundary that includes the partially matched shape is always shorter than the length of the boundary of the prior shape in its matched scale. Therefore, the shape similarity result only needs to be computed between C_{V,σ_i} and a subset of curvatures in S .

4 Stochastic Region Grouping with Shape Prior

This section follows the mathematical framework proposed by Tu et al. [17] and derives a Bayesian formulation for grouping image regions with the introduced shape prior. We use a “region adjacency graph” [1] that contains a vertex for each atomic region of the oversegmented image and an edge e_{ij} between vertices v_i and v_j if the regions represented by v_i and v_j are adjacent in the image (Fig. 3(a)–(c)). Given a region adjacency graph, an image segmentation W is defined as a random variable whose assignments correspond to segmentation states during the region grouping process. In particular, in a given segmentation state,

$$W = ((V_1, \theta_1), (V_2, \theta_2), \dots, (V_n, \theta_n)), \quad (3)$$

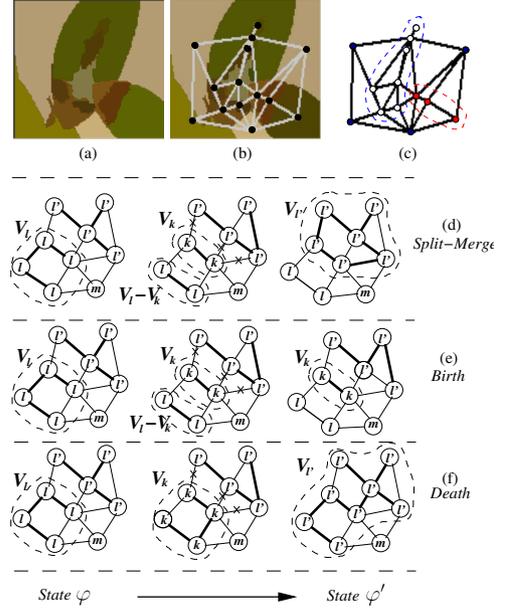


Figure 3: Region Adjacency Graph and Grouping Operations: (a) Oversegmented image of leaves occluding each other. (b) Region adjacency graph with vertices placed on the centroid of each atomic region. (c) Region grouping result where the vertices belonging to the two leaves are marked in red and white, respectively. (d)–(f) Three types of region grouping operations for the segmentation state transition from φ to φ' . Vertices with the same label belong to the same region group and are connected by “turned-on” (thick) edges. During the state transition (middle column), a subgraph V_k of a group V_i is chosen, which either becomes a new region (e), or merges into a neighboring group V'_i (in (d) $V_k \subset V_i$ and in (f) $V_k = V_i$). The set of edges between groups V_k and V'_i before the merging operation are defined as $Cut(V_k, V'_i - V_k)$ (marked with crosses).

where V_1, \dots, V_n are mutually disjoint region groups or subgraphs, $V_i \cap V_j = \emptyset$, each group V_i is generated by organizing a number of atomic regions, $V_i = \{v_k^i\}$ via a series of region grouping (graph partition) operations [1] (Fig. 3(d)–(f)), and parameter θ_i defines an image model that can be learned in advance. The objective of the region grouping process is to achieve a good segmentation that not only provides results consistent with human perceptual groupings but also yields object recognition, i.e., identifies region groups that correspond to the objects of interest. This grouping process should be influenced by image and shape constraints simultaneously. If we assume V_1, \dots, V_n are mutually independent random variables, given an observed image I and a shape prior S , the posterior probability for the segmentation W is

$$p(W|I, S) \propto p(I|S, W)p(S|W)p(W) \quad (4)$$

$$\propto \left[\prod p(I_{V_i}|V_i, S) \right] \left[\prod p(S|V_i) \right] p(W) \quad (5)$$

$$\propto \left[\prod p(I_{V_i}|\theta_i, S) \right] \left[\prod p(S|C_{V_i}) \right] p(W), \quad (6)$$

where I_{V_i} represents image patch associated with V_i , and C_{V_i} stores the boundary curvatures of the region V_i . The

likelihood
$$p(I_{V_i}|\theta_i, S) \propto \exp(-D(I_{V_i}, \theta_i)) \quad (7)$$

is computed based on the function $D(I_{V_i}, \theta_i)$ that measures the compatability of the observed image data I_{V_i} with the objects' predefined appearance models θ_i . These may be Gaussian or histogram-based color or texture models that can be learned in advance. The shape posterior is

$$p(S|C_{V_i}) \propto \exp(-(1 - M(V_i, S))), \quad (8)$$

where $M(V_i, S)$ is used to measure the shape similarity between the current region V_i and the shape prior S (Eq. 2). As suggested previously [17], the number n of region groups and the size $|V_i|$ and boundary smoothness of each region group are taken into account in defining the prior probability

$$p(W) \propto \exp(-c_1 n - c_2 \sum_i^n |V_i|^\tau - c_3 \sum_i^n |C_{V_i}|), \quad (9)$$

where $|C_{V_i}|$ is the sum of curvature magnitudes along V_i 's boundary, and c_1, c_2, c_3 and τ are some constants. Intuitively, a segmentation is likely to include a small number of large regions with smooth boundaries.

A solution of the image segmentation problem is obtained by simulating the posterior probability $p(W|I, S)$ via a Markov chain and finding the segmentation W that maximizes it. The Markov chain can be realized by a Metropolis-Hastings mechanism [19, 17]. Given the solution space $\Omega = \{\varphi | \varphi \text{ is a possible state of } W\}$ of the segmentation problem, we define $\varphi, \varphi' \in \Omega$ to be the two configurations of W that respectively correspond to the segmentation results before and after a region grouping operation (Fig. 3). The probability $q(\varphi \rightarrow \varphi') = p(\varphi'|\varphi, I, S)$ indicates how likely it is to transfer from the current state φ to the next state φ' . When a state transition from φ to φ' is accepted by

$$\alpha(\varphi \rightarrow \varphi') = \min(1, \frac{q(\varphi' \rightarrow \varphi)}{q(\varphi \rightarrow \varphi')} \cdot \frac{p(\varphi'|I, S)}{p(\varphi|I, S)}), \quad (10)$$

the Metropolis-Hastings method guarantees that the Markov chain will converge to $p(W|I, S)$ as its stationary distribution. Therefore, given the definition of $p(W|I, S)$, there is a large chance a good segmentation can be achieved after many sampling iterations.

4.1 Swendsen-Wang Cut Algorithm

One major disadvantage for most MCMC methods is that, a long simulation process is usually required for convergence. The recent Swendsen-Wang Cut (SWC) method [1] generalized a well accepted cluster sampling algorithm [15] for solving a graph partition problem. This method allows a large sampling move between very different graph configurations, thus providing fast simulation and optimization. For the current problem, we apply the SWC-2 algorithm (journal preprint of [1]) combined with other modifications to sample the different segmentation configurations and perform the region grouping operations, so that

MOSAICSHAPE (I, S, θ_I, P)

// I : image; S : prior shape; θ_I : image model; P : other parameters;

1. Generate oversegmented atomic regions for I .
2. **Compute boundary curvature for each atomic region.**
3. **Compute band probabilities b_{ij} between adjacent atomic regions.**
// Sample $p(W|I, S)$ by transition move $\varphi \rightarrow \varphi'$ (Fig. 3)
4. For the current segmentation φ ,
 - 4.1. **Choose an unmarked atomic region v_k randomly and record its parent region group as V_l .**
 - 4.2. Turn on the edge e_{ij} with the band probability b_{ij} inside V_l .
 - 4.3. Given the turned-on e_{ij} inside V_l , find the connected region component for v_k and record it as V_k .
5. **Merge V_k into some adjacent region group $V_{l'}$ with probability $q(l'|V_k, \varphi, I, S)$ and record this new state as φ' .**
6. **Generate V_Σ by the shape registration with the new $V_{l'}$.**
7. **If $V_{l'}$ or V_Σ satisfies given shape similarity criteria w.r.t S , record $V_{l'}$ or V_Σ as one of *Recognized Objects* and identify its atomic regions as marked.**
8. Accept the new state, e.g., $\varphi = \varphi'$, with the probability $\alpha(\varphi \rightarrow \varphi')$.
9. Repeat Step 2-9 until some convergence criterion or the expected number of the objects is achieved.

Return all identified region groups as *Recognized Objects*.

an ideal segmentation result can be achieved efficiently. The main steps of the new algorithm *MosaicShape* are summarized below, where the modified parts will be described later (accentuated in bold). The main conclusion from previous work [1] was that when:

$$\frac{q(\varphi' \rightarrow \varphi)}{q(\varphi \rightarrow \varphi')} = \frac{\prod_{e \in \text{Cut}(V_k, V_{l'} - V_k)} (1 - b_e)}{\prod_{e \in \text{Cut}(V_k, V_l - V_k)} (1 - b_e)} \cdot \frac{q(l'|V_k, \varphi', I, S)}{q(l'|V_k, \varphi, I, S)} \quad (11)$$

was defined for $\alpha(\varphi \rightarrow \varphi')$ (Eq. 10), the algorithmic process was ergodic, aperiodic and had $p(W|I, S)$ as its stationary distribution, where b_{ij} (or b_e) is the band probability that determines how likely a pair of adjacent atomic regions should be grouped together, $q(l'|V_k, \varphi, I, S)$ is the transition probability that defines how likely it is for the region group V_k to be merged with its neighboring region group $V_{l'}$, and $\text{Cut}(V_k, V_{l'} - V_k) = \{e_{ij} | v_i \in V_k, v_j \in (V_{l'} - V_k)\}$ was defined as the set of region edges between region group V_k and $V_{l'} - V_k$, as shown in Fig 3. As to be described next, once b_{ij} (or b_e) and $q(l'|V_k, \varphi, I, S)$ are defined, the acceptance probability $\alpha(\varphi \rightarrow \varphi')$ can be computed directly given Eq. 11 and $p(W|I, S)$ defined in Eq. 4.

4.2 Band and Transition Probabilities

A band probability $b_{ij} = p(e_{ij} = \text{"on"} | v_i, v_j, I, S)$ is introduced for the edge e_{ij} between two adjacent atomic regions v_i, v_j . In the previous work [1], it is defined to be large when the image information of two regions is compatible. For the current problem, b_{ij} should also be large if the shape of the merged region is more similar to the shape prior than either v_i or v_j . It is therefore defined as the product

$$b_{ij} \propto p(I|v_i, v_j) p(S|v_i, v_j) \propto e^{-\mathcal{M}_I(I_{v_i}, I_{v_j})} \cdot e^{-\mathcal{M}_S(v_i, v_j, S)} \quad (12)$$

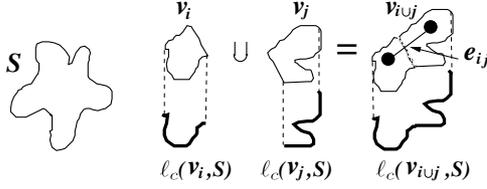


Figure 4: **Shape-Based Band Probability:** The longest partially matched boundaries of the image regions are shown in bold.

where the shape measure $\mathcal{M}_S(v_i, v_j, S) =$

$$\eta \frac{M(v_{i \cup j}, S)}{\max(M(v_i, S), M(v_j, S)) + M(v_{i \cup j}, S)},$$

$$\eta = \max\{1, \ell_c(v_{i \cup j}, S) / \max\{\ell_c(v_i, S), \ell_c(v_j, S)\}\},$$

and the image measure

$$\mathcal{M}_I(I_{v_i}, I_{v_j}) \propto \sum_k \frac{(H_i(k) - m(k))^2}{m(k)},$$

where $H_i(k)$ and $H_j(k)$ are the regions' k th bins of their respective intensity histograms, and $m(k)$ is their average $(H_i(k) + H_j(k))/2$. Many dissimilarity measurements could be applied to compute the distance between two histograms [13]; in the current implementation, a χ^2 statistics distance is computed.

As illustrated in Fig. 4, the new region $v_{i \cup j}$ is introduced by merging v_i and v_j , and $\ell_c(v_{i \cup j}, S)$ represents the length of the longest matched subsequence between the curvature sequence of $v_{i \cup j}$ and those sequences stored in S . On one hand, the shape similarity defined by $M(v_{i \cup j}, S)$ in algorithm *PSM* (Section 3.1) is normalized by the boundary length of $v_{i \cup j}$ so that a large region with the comparable length of the matched boundary is penalized. On the other hand, the scale factor η plays the role of encouraging the existence of a large region. In particular, η could be greater than 1 when $v_{i \cup j}$ was matched with the shape prior in a large scale, while $M(v_{i \cup j}, S)$ is smaller than $\max(M(v_i, S), M(v_j, S))$.

As shown in Fig. 3, the neighboring region groups for the chosen V_k can be represented by

$$\{V_1, V_2, \dots, V_l - V_k, \dots, V_n, \emptyset\} \quad (13)$$

and indexed from 1 to $n + 1$, where $V_l - V_k$ represents the remaining region group after V_k was split from V_l . The transition or ‘‘proposal’’ probability $q(l'|V_k, \varphi, I, S)$ [1] defines how likely a merge of V_k with a region group $V_{l'}$ is among all candidate groups. It is defined similar to Eq. 12 as

$$q(l'|V_k, \varphi, I, S) = p(V_k \text{ is merged with } V_{l'}) \quad (14)$$

$$\propto p(I|V_k, V_{l'}) p(S|V_k, V_{l'})$$

and normalized by $\sum_{i=1}^{n+1} q(i|V_k, \varphi, I, S)$.

Since the band (Eq. 12) and transition move (Eq. 14) probabilities characterize the dominant properties modeled by the posterior probability $p(W|I, S)$ (Eq. 4) well, the

transition move $\varphi \rightarrow \varphi'$ is expected to be accepted with a high probability such that the designed Markov chain will quickly converge to the desired solution.

4.3 Sampling Jump by Shape Registration

The shape matching process (Section 3.1) can provide strong constraints on the correspondence between the matched image region and the shape prior. This allows the simulation process to realize a large sampling jump. Given a region group V and the shape prior S , the sets of the corresponding points on their matched boundaries are recorded as P_V and P_S , respectively. A pair of translation and rotation parameters (t, α) are calculated by the least-square method for registering P_S with P_V . The prior shape S in the matched scale can then be transformed onto the image and noted as S_T . A new region group V_Σ can be generated by merging all atomic regions inside S_T :

$$V_\Sigma = \{v_i | v_i \in S_T, \text{ for all } i\}, \quad (15)$$

where the operation \in judges if the centroid of v_i is inside S_T . If the size of V_Σ is comparable to the shape prior in the matched scale, the posterior probabilities (Eqs. 7 and 8) are then computed for V_Σ . Region group V_Σ is recorded as the recognized object when the obtained posterior probability is larger than a given threshold, depending on what degree of occlusion is allowed for the objects to be retrieved. In the current implementation, we embedded this operation within each transition move and checked if any matched object could be recognized, while continuing with the Metropolis-Hasting sampling in its usual way.

5 Some Implementation Issues

The *MOSAICSHAPE* algorithm first applied the mean shift method [4] to generate an oversegmented image (Step 1), in which a bandwidth value and the minimum region size were chosen so that the number of generated oversegmented regions was moderate. To compute the curvature values for each region contour, the Fast Fourier Transformation (FFT) and inverse FFT were applied to speed up the convolution operations in Eq. 1. The evaluation of probabilities defined in Eqs. 4, 12 and 14 was implemented by a regularization framework, where a small weight (0.2) was assigned for the relevant image probability, and a large weight (0.6) for the relevant shape probability. Moreover, to speed up the overall simulation process, for operations requiring expensive computations, such as curvature calculation, partial shape matching, or color histogram construction, results were only computed once for a newly generated region group, then stored into a sorted linked list indexed by the labels of the atomic regions within the current region group. Afterwards, when the same region group was re-

visited, these results could be accessed in time complexity $O(\log(n))$.

6 Experiments

The proposed method was tested on both synthetic and real images. In the synthetic experiments, an irregular “star shape” was first created as the shape prior S (Fig. 5a) and stored in a multi-scale curvature form, where the scale range was from 0.75 to 1.25 compared to the mean size, and the widths for the chosen Gaussian kernels were 1, 2 and 4. Synthetic images were then created by randomly placing several mosaic “star” objects onto real images with a lot of clutter in the background. Segmentation results were obtained where star objects of different rotations and sizes occurred with complete (Fig. 5) or partial shape (Fig. 6). In the latter case, the desired partial objects were retrieved by specifying an acceptable shape similarity threshold (0.7 in this experiment).

To test the MOSAICSHAPE algorithm on real images, an ellipse shape prior was first learned from a set of 12 leaves. The goal was to detect leaves with color patches caused by autumn. Different color distributions were observed on the surfaces of these leaves, which yielded a number of small atomic regions in the oversegmented images. To cover the shape variability among the training samples, the registration parameters (t, α) were first computed between the image regions and the learned average shape. These parameters were then applied for registering each shape sample in the training set onto the image. Experimental results for detecting the objects with complete or partially matched shapes are shown in Fig. 7 and 8.

The overall segmentation process took 30-90 s for synthetic and 150-250 s for real images of size about 200-250 by 200-250. As can be seen in most experiments, our method provided satisfactory results for retrieving the shape from the images, where the stochastic simulation process usually started as a slow annealing process and recognized the target objects by making a large sampling jump once a good partial matching criterion was met. In some situations, difficulties inherent in the original partial shape matching problem may lead to ambiguous situations (Fig. 8 2(c-d)) by which the convergence process was slowed down.

7 Discussion and Conclusion

The paper proposed a novel framework for object segmentation and recognition. Its main contribution was to integrate the decomposed shape constraints into a bottom-up image segmentation process using partial shape matching. By this means, the segmentation and recognition of multiple occluded objects can be achieved simultaneously.

The current method can be improved in several aspects. First, the results of the partial shape matching could be am-

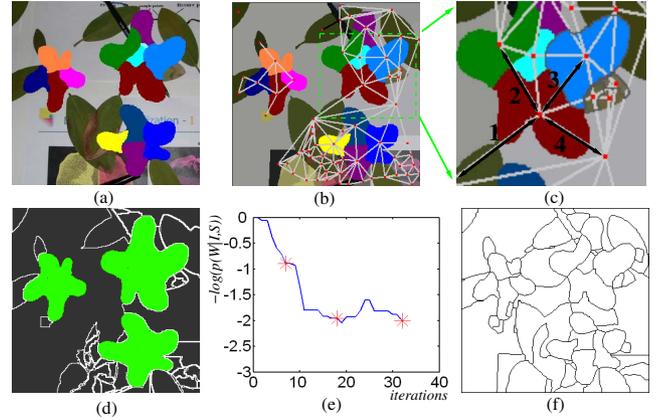


Figure 5: (a) Input image. (b) Oversegmented image [4] with region adjacency graph. (c) Close-up view for some local region. The respective band probabilities for edges 1–4 are 0.24, 0.68, 0.67 and 0.39 (Eq. 4.2). (d) Segmentation result (green). (e) Posterior energy during the simulation. The sampling states (red stars) at which the target objects were identified. (f) Segmentation result by DDMCMC method [17].

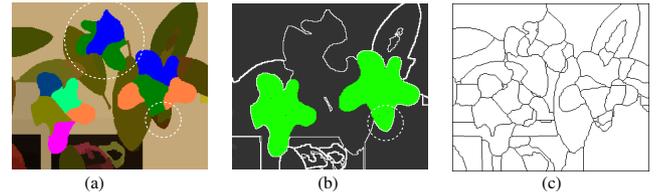


Figure 6: (a) Oversegmented image [4]. Object in large white circle has partial shape matching similarity 0.64. (b) Our result. Atomic regions in small white circle were included in a matched object due to a large sampling jump. (c) Segmentation result by DDMCMC method [17].

biguous during the region grouping process (Fig. 8 2(c-d)). These ambiguities can be decreased when more sub-regions are merged. Adopting better shape similarity measurements will produce more accurate shape matching results and lead to a faster shape recognition process. For instance, a dynamic programming method [11] may be applied to allow the matching of distorted shapes. Second, the shape prior used in the current method is limited to the class of 2D planar objects whose shape can be easily represented by their boundary curvatures. In the future, we plan to perform extra experiments on more complicated real-world objects. However, the problem of defining the shape for objects with non-rigid appearance, e.g., human or animal, or for objects in higher dimensions (3D), is ambiguous or very challenging in itself. Moreover, the solution of partial shape matching will become elusive when the intra-shape variations needs to be modeled, for instance, by Principle Component Analysis (PCA) method. Finally, integrating comprehensive image models [17, 12] or feature-based template matching [2] into the current system may help capture additional variabilities of the object appearance and improve segmentation

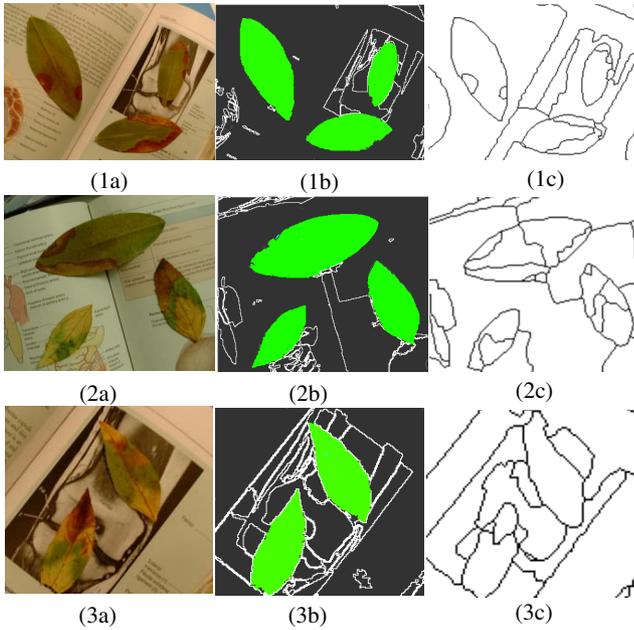


Figure 7: 1st column: Input images; 2nd column: Our results; 3rd column: Results by DDMCMC [17].

results.

Acknowledgment

The authors thank Dr. Adrian Barbu at UCLA for helpful discussions on Swendsen-Wang Cut and providing the segmentation results for test images by the DDMCMC program, developed by Z.W. Tu and S.C. Zhu [17]. Financial support by NSF (grants 0093367, 0329009, 0202067, 0326483) is also gratefully acknowledged.

References

- [1] A. Barbu and S. Zhu. Graph partition by Swendsen-Wang cuts. In *ICCV*, pages 320–327, Nice, France, Oct. 2003. A preprint of journal version can be found at <http://www.cs.ucla.edu/~abarbu/>.
- [2] E. Borenstein, E. Sharon, and S. Ullman. Combining top-down and bottom-up segmentation. In *IEEE Workshop on POCV*, Washington DC, USA, 2004.
- [3] Y. Chen, H. Tagare, and et al. Using prior shapes in geometric active contours in a variational framework. *IJCV*, 50:315–328, 2002.
- [4] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *PAMI*, 24(5):603–619, 2002.
- [5] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1:321–331, 1987.
- [6] J. Kaufhold and A. Hoogs. Learning to segment images using region-based perceptual features. In *CVPR*, pages 954–961, Washington DC, USA, 2004.
- [7] M. E. Leventon, W. E. L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *CVPR, Vol 1*, pages 316–323, Hilton Head, SC, USA, June 2000.

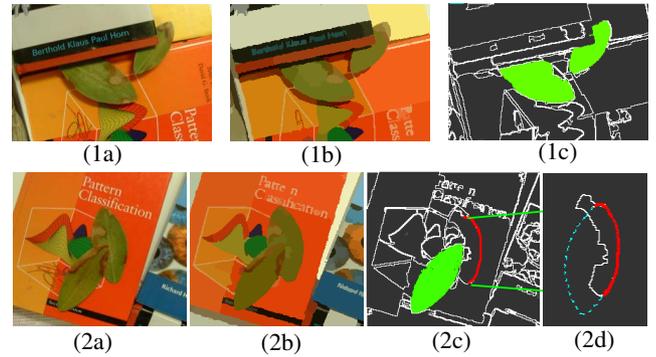


Figure 8: 1(a)-2(a): Input images; 1(b)-2(b): Oversegmented images [4]; 1(c)-2(c): Our results. The matching ambiguity inherent from the partial shape match could directly affect the process of recognizing the object of interest. In 2(d), the occluded leaf with a partially matched boundary (in red) could lead into an unexpected shape registration result (in blue dash).

- [8] S. Loncaric. A survey of shape analysis techniques. *Pattern Recognition*, 31(8):983–1001, 1988.
- [9] R. Malladi, J. Sethian, and B. Vemuri. Shape modeling with front propagation: A level set approach. *PAMI*, 17:158–175, 1995.
- [10] F. Mokhtarian. Silhouette-based isolated object recognition through curvature scale space. *PAMI*, 17(5):539–544, 1995.
- [11] E. Petrakis, A. Diplaros, and E. Millos. Matching and retrieval of distorted and occluded shapes using dynamic programming. *PAMI*, 24(11):1501–1516, 2002.
- [12] X. Ren and J. Malik. Learning a classification model for segmentation. In *ICCV*, pages 10–17, Nice, France, Oct. 2003.
- [13] Y. Rubner, C. Tomasi, and L. Guibas. The earth mover’s distance as a metric for image retrieval. *IJCV*, 40:99–121, 2000.
- [14] S. Sclaroff and L. Liu. Deformable shape detection and description via model-based region grouping. *PAMI*, 23(5):475–489, 2001.
- [15] R. Swendsen and J. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Phys Rev Lett*, 58(2):86–88, 1987.
- [16] Z. W. Tu, X. R. Chen, A. L. Yuille, and S. C. Zhu. Image parsing: unifying segmentation, detection, and recognition. In *ICCV*, pages 18–25, Nice, France, Oct. 2003.
- [17] Z. W. Tu and S. C. Zhu. Image segmentation by data-driven Markov Chain Monte Carlo. *PAMI*, 23(5):657–673, 2002.
- [18] R. C. Veltkamp. Shape matching: Similarity measures and algorithms. In *Proceedings of the International Conference on Shape Modeling and Applications*, pages 188–199, Genova, Italy, 2001.
- [19] G. Winkler. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*. Springer Verlag, 1995.