Boston University Computer Science Technical Report No. 2005-008, Feb. 2005. To appear in Proc. CVPR, 2005.

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
†{jingbinw,betke}@cs.bu.edu; ‡erdan@seas.upenn.edu

Abstract

A novel 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 solved by a stochastic Markov Chain Monte Carlo (MCMC) mechanism. By this means, object segmentation and recognition are accomplished simultaneously. 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 parallel 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

Successful segmentation or retrieval of objects of interest from images is of crucial importance for a wide range of applications. A successful method should provide a good approximation to the optimal segmentation solution under various situations. However, the appearance of an object in real images can be affected by many factors, such as different color/texture distributions of the object appearance or various illumination conditions. It becomes particularly challenging when the object is present in the foreground with other objects as clutter. One solution strategy is to incorporate high-level prior knowledge, such as a shape prior, into an existing image segmentation method.

Contour-based segmentation methods, e.g., active contours [6] or level set methods [10], explicitly or implicitly deform an active contour to capture the boundary of the object. In previous works by Leventon et al. [8] and Chen et al. [3], learned shape priors were introduced to constrain the 2D (3D) contour (surface) deformations so that target objects with theses predefined boundary shapes could be extracted from the cluttered background. However, the original limitations inherent in contour-based methods, such as the well-known initialization and local minima problems (Fig. 1a), still remained.

(a) (b) (c)

Figure 1: (a): Contour-based segmentation result by the traditional active contour method [6], where the initial contour was shown as the link of blue knots and the final contour was represented as the link of red knots. (b): Region-based segmentation result by data-driven MCMC [18]. (c): Segmentation result by the proposed method.

Compared to the contour-based methods, recent region-based methods [18, 13, 7] have the following advantages. First, region-based methods are bottom up and data driven. They do not require an initialization step in general, and can approximate the globally optimal solution in many cases. Second, different types of high-level prior knowledge, such as color/texture models [18, 13], boundary continuity hypotheses [18, 13], or perceptual measurements [7], can be incorporated into the bottom-up segmentation process. However, since this prior knowledge typically cannot capture the large variabilities of an object's appearance or shape that occur in practice, application of the above region-based methods was mainly restricted to generate perceptual groupings or visually pleasing segmentation results (Fig. 1b).

This paper proposes a novel method to incorporate prior knowledge of shape into a bottom-up region-based segmentation process for segmenting and recognizing objects of the interest in images (Fig. 1c). Our main contributions are summarized below:

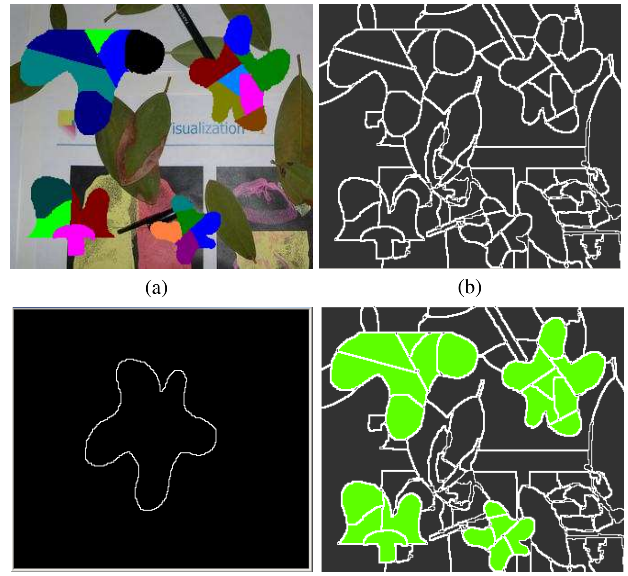
- Given prior knowledge of shape, the proposed method identifies target objects in images by grouping oversegmented image regions via a stochastic Markov Chain Monte Carlo (MCMC) mechanism[1]. By this means, segmentation and recognition of multiple occluded object are accomplished simultaneously.
- During 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 a parallel algorithm, called *Swendsen-Wang Cut* (SWC) algorithm [1], to the current problem and providing new implementation improvements.

The work most relevant to the current method was proposed by Sclaroff et al. [15]. They used a deformable shape template as a shape constraint for grouping image regions. Their method was deterministic and assumed the correct initial condition was guaranteed in most cases. Moreover, object segmentation via deformable shape templates [21, 5] may not capture the large variability of the shape within the given class. As a result, large occlusions cannot be handled [15]. Recent work by Tu et al. [17] pushed in the direction of accomplishing object segmentation and recognition simultaneously within a unified framework [18]. Another idea of combining the top-down and bottom-up segmentation was demonstrated recently by Borenstein et al. [2] for segmenting the foreground objects from images. These systems required the feature based prior models to be carefully constructed through a time-consuming learning process. Therefore, they were of limited use for retrieving objects of large variability in appearance or of arbitrary shapes from images. Moreover, these methods did not handle occluded objects explicitly.

2. Problem Definition and Method Overview

Given an input color image (Fig. 2a), its oversegmented image (Fig. 2b) can be obtained by some existing method, for instance, a mean shift method [4]. However, because an object of interest may be partitioned into multiple atomic regions, such an oversegmented image cannot directly provide meaningful object interpretations. When a shape prior is introduced (Fig. 2c), a meaningful segmentation can be achieved (Fig. 2d), where the objects of the interest in the image are identified and recognized.

The segmentation problem to be solved is, given a shape prior represented in some form, how to group the atomic regions in the image such that the objects similar to the shape prior can be correctly identified. Finding the globally optimal solution for such a problem is NP-hard [15]. On one hand, a deterministic method [15] does not guarantee an optimal solution in general. For instance, because an image region may only provide a partial hypothesis for the shape similarity measurement, a best matched partial region boundary does not necessarily imply that it will be matched with the shape prior in the final optimal solution. On the other hand, an exhaustive top-down shape matching process could be extremely slow because the shape of the objects in the image can be translation, rotation, and scale invariant. Therefore, we apply a stochastic mechanism as a compromise between the above two strategies in the proposed method.



(a) (b) (c) (d)

Figure 2: (a): Input color image. (b): Oversegmented image. (c): Shape prior. (d): Recognized objects of interest, including either complete or partial shape information.

Section 3 describes how a shape prior is represented and used for measuring the shape similarity between an image region and the prior shape by performing a partial shape matching. In Section 4, a stochastic MCMC mechanism is formulated as the proposed solution for the given segmentation problem. A *Swendsen-Wang Cut* algorithm is applied to carry out the stochastic mechanism and speed up the overall computational process(Section 4.1), within which, the probabilistic region grouping operations are carefully designed by taking both the image and shape constraints into account (Section 4.2). During the simulating process, a large sampling jump for fast convergence can be obtained by a simple shape registration (Section 4.3). Details about system implementation are given in Section 5.

1 The mathematical background for MCMC theory was given in [20].

3. Shape Prior and Multi-Scale Curvature Representation

Defining a good representation of a shape is a challenging problem in itself [9, 19]. For the problem at hand, the representation of the shape prior must be translation, rotation, and scale invariant. In the proposed method, we therefore use boundary curvatures to define the shape of an object. Given the 2D closed contour of an object, parameterized by the arc length parameter as $\{(x(u), y(u))\,|u \in [0, 1]\}$, a smoothed version of boundary curvature can be calculated [11] by convolving the curve with different sizes of Gaussian kernels to yield $(x(u, \sigma), y(u, \sigma))$, where σ is the width of the kernel, and then computing

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

To achieve a scale invariant representation, Ref. [11] normalized the curvature values by the length of the contour, which was however problematic for handling the partial shape matching problem described below. The proposed method therefore instead precomputes a set of boundary curvatures for the object present in different scales. In particular, a shape prior S is defined as

$$S = \{C^*_{i,\sigma_j}, i \in [1, \ldots, m], j \in [1, \ldots, n]\}, \quad (2)$$

which consists of sequences of curvature values C^*_{i,σ_j} along the boundary of the shape prior, smoothed by a Gaussian kernel of width σ_j in the ith level of scale. The total number of such curvature sequences are $m \times n$, for m different levels of scale and n different sizes of Gaussian kernels.

3.1. Partial Shape Matching Problem

The boundaries of oversegmented image regions typically only partially match the contour of a given shape prior (Fig. 2b). In order to apply the information on prior shape to group the oversegmented image regions, partial shape similarity between the image regions and the given shape prior needs to be measured. To our knowledge, finding a general solution for identifying the matches of different parts of the shapes is still an unsolved problem [19, 12]. In the problem at hand, we assume the partially matched portions of the boundary of the same object are connected and only small "gaps" are allowed between them. Given an input object V (e.g., an image region) and a shape prior S, two boundary curvature sequences C_V for V and C^* for S can be computed by Eq. 1. A revised 1D correlation process is implemented (Algorithm 1) to identify the longest subsequence that matches in both C and C^*. The length ℓ of this subsequence is normalized by the boundary length of the input object. The second iteration, repeating $2\,\ell(C_V)$ times, allows the matching to start from any position on the region boundary due to its cyclic representation. Threshold E controls the gap size allowed in the final matched sequence, e.g.,, $E \in [0.1, 0.15]$ in the current implementation.

Algorithm 1 : PARTIALSHAPECORRELATION(Curvatures C_V of Input Object, Curvatures C^* of Prior Shape, Curvature Similarity Threshold T, Allowed Gap Size E)

```
Maxhits = 0
for i = 1 to ℓ(C*) do
    Hits = Unhits = 0
    for j = 1 to 2 ℓ(C_V) do
        if j > ℓ(C_V)  then  m = j − ℓ(C_V)
        if (i + j) >= ℓ(C*)  then  n = i + j − ℓ(C*)
        if |C(m) − C*(n)| < T  then  Hits++
        else  Unhits++
        if Unhits / Hits > E  then
            if Hits > Maxhits  then  Maxhits = Hits
            UnHits = Hits = 0
Normhits = Maxhits /ℓ(C_V)
Return Normhits
```

To measure the shape similarity $M(V, S)$ between V and S, Algorithm 1 requires to be performed for all size and smoothness levels

$$M(V, S) = \max\{d_{i,j,k} = \mathrm{psc}(C_{V,\sigma_i}, C^*_{j,\sigma_k}), \forall i, j, k\} \quad (3)$$

where psc() is the partial shape correlation function computed by Algorithm 1, and 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

In this section, we follow the mathematical framework proposed by Tu et al. [18, 1] and derive a Bayesian formulation for grouping image regions with the introduced shape prior. We use a "region adjacency graph," which contains a vertex for each atomic region of the oversegmented image and an edge e_{ij} between vertices v_i and v_j in the graph if the regions represented by v_i and v_j are adjacent in the image, i.e., share a boundary (Fig. 3(a)–(c)). During the segmentation process, the atomic regions may be labeled as different region groups via a series of dynamic region grouping (graph partition) operations [1] (Fig. 3(d)–(f)). The objective of the current segmentation task is to correctly group the atomic regions so that the region groups satisfying the given image and shape constraints can be found and identified as the objects of interest. Given a region adjacency graph, an image segmentation is represented by:

$$W = (n, (V_1, \theta_{I1}), (V_2, \theta_{I2}), \ldots, (V_n, \theta_{In})) \quad (4)$$

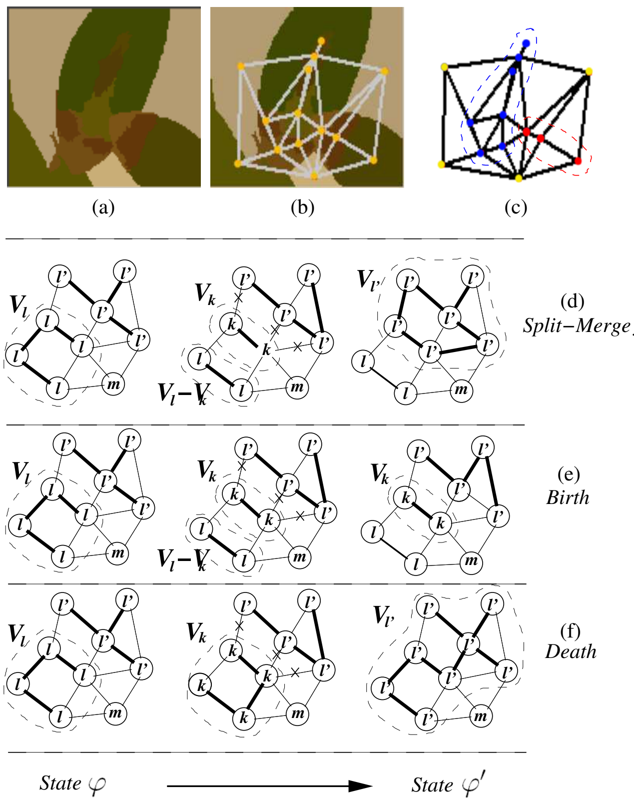


Figure 3: **Region Adjacency Graph and Region Grouping Operations:** (a) An oversegmented image of leaves occluding each other. (b) Image with region adjacency graph and graph vertices placed on the centroid of each atomic region. (c) A region grouping result where the vertices belonging to the two leaves are marked in blue and red, respectively. (d)-(f) Three types of region grouping operations are defined for the segmentation state transition from φ to φ', where vertices with the same label are connected by "turned-on" edges (thicker edges) and marked as the same region group. During the state transition, a subgraph V_k is chosen in V_l, then it could be merged into one of its neighboring groups ((d) for $V_k \subset V_l$ and (f) for $V_k = V_l$), or become a new region group (e). A set of region edges between V_k and V_l' before the merging operation are defined as $Cut(V_k, V_{l'} - V_k)$ and marked with crosses in the middle column.

as a random variable, whose different assignments correspond to different segmentation states during the region grouping process. For instance, $V_1, \ldots, V_n$ are n region groups or subgraphs in some state, and each V_i may include a number of atomic regions, such as $V_i = \{v_1^i, \ldots, v_n^i\}$ and $V_i \cap V_j = \varnothing$. Parameter θ_{Ii} summarizes a predefined image model for each image region and can be learned in advance. If we assume $V_1, \ldots, V_n$ are mutually independent, given an observed image I and a shape prior S defined as previous, a posterior probability for the segmentation W is:

$$p(W|I,S) \propto p(I|S,W)p(S|W)p(W) \quad (5)$$

$$\propto [\prod_{i=1}^{n} p(I_{V_i}|V_i,S)][\prod_{i=1}^{n} p(S|V_i)]p(W) \quad (6)$$

$$\propto [\prod_{i=1}^{n} p(I_{V_i}|\theta_{Ii},S)][\prod_{i=1}^{n} p(S|C_{V_i})]p(W) \quad (7)$$

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

$$p(I_{V_i}|\theta_{Ii},S) \propto \exp(-D(I_{V_i},\theta_{Ii})) \quad (8)$$

$$p(S|C_{V_i}) \propto \exp\left(-(1 - M(V_i,S))\right) \quad (9)$$

$$p(W) \propto \exp\left(-c_1 n - c_2 \Sigma_i^n |V_i|^\tau - c_3 \Sigma_i^n |C_{V_i}|\right), \quad (10)$$

respectively. Here, the discriminative model $D(I_{V_i}, \theta_{Ii})$ measures the compatibility of the observed image data I_{V_i} and the predefined appearance models θ_{Ii} for the target objects. It can contain one of two types of information: a color model defined as a Gaussian $G(\mu, \sigma)$ with specified mean μ and variance σ or a texture model as an n-bin histogram $H(n, h_1, \ldots, h_n)$ on different intensity levels $(h_1, \ldots, h_n)$. Both models can be learned in advance depending on the appearance of the object of interest. Similarity $M(V_i, S)$ is defined in Eq. 3 and measures the shape similarity between the current region V_i and the shape prior S. As suggested by the previous statistical study [18], the number n of region groups, the size $|V_i|$ of each region group and the boundary smoothness $|C_{V_i}|$ of each region are taken into account for defining the prior probability $p(W)$ in Eq. 10, where $|C_{V_i}|$ represents the summation of curvature magnitudes along the region boundary. Intuitively, a segmentation is likely to include a small number of large regions with smooth boundaries.

The solution for the above segmentation problem is approximated by simulating the posterior probability $p(W|I,S)$ via a Markov chain. It can be realized by a Metropolis-Hastings mechanism [20]. For our problem, sampling the segmentation states of W corresponds to a series of region grouping operations. 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 $q(\varphi \to \varphi') = p(\varphi'|\varphi, I, S)$ indicates how likely it is for the current state φ to transfer to the next state φ'. When a proposed move from φ to φ' is accepted by

$$\alpha(\varphi \to \varphi') = \min(1, \frac{q(\varphi' \to \varphi)}{q(\varphi \to \varphi')} \cdot \frac{p(\varphi'|I,S)}{p(\varphi|I,S)}), \quad (11)$$

the Metropolis-Hastings method guarantees that the above 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, during each sampling iteration, only a small local sampling move is allowed and the adjacent states are generally similar. Therefore, for convergence, a long simulation process is usually required. The recently developed Swendsen-Wang Cut (SWC) method [1] generalized a well accepted stochastic parallel algorithm [16] and was applied to solve 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 an ideal segmentation result can be achieved efficiently. The main steps of the new algorithm MOSAICSHAPE are summarized below, where the modified parts are accentuated in bold and will be described further.

Algorithm 2 : **MOSAICSHAPE** (Image I, Prior Shape S, Image Model θ_I, Other Segmentation Parameters P)
Return value: List of Retrieved Objects;

---

1. Generate over-segmentation results for I.
2. **Compute boundary curvature for each atomic region.**
3. **Compute band probabilities** b_{ij} **between all pairs of adjacent atomic regions.**
   // Sample $p(W|I,S)$ by proposed move $\varphi \rightarrow \varphi'(Fig.\,3)$
4. For the current segmentation φ,
   - 4.1. **Randomly choose an unmarked atomic region** v_k **and record its parent region group as** V_l**.**
   - 4.2. Turn on the edge e_{ij} with the band probability b_{ij} for all pairs of adjacent atomic regions inside the group V_l.
   - 4.3. Follow the turned-on edges connected with the chosen v_k, find the connected region component inside V_l, and record it as V_k.
5. **Merge** V_k **with its adjacent region group** $V_{l'}$ **with probability** $q(l'|V_k,\varphi,I,S)$ **and record this new state as** φ'**.**
6. **Check if region group** $V_{l'}$ **is similar enough to** S**, or, if a large sampling jump is allowed, record the matched region group into a List of Retrieved Objects and mark its atomic regions.**
7. Accept the new state $\varphi = \varphi'$ with the probability with the probability $\alpha(\varphi \rightarrow \varphi')$.
8. Repeat from Step 2 until the convergence criterion is met or the expected number of the objects is retrieved.

To apply the MOSAICSHAPE algorithm, three probabilities, namely, b_{ij}(or b_e), $q(l'|V_k,\varphi,I,S)$ and $\alpha(\varphi \rightarrow \varphi')$, must be defined in advance. The main conclusion from previous work [1] was that when:

$$\frac{q(\varphi' \rightarrow \varphi)}{q(\varphi \rightarrow \varphi')} = \frac{\prod_{e \in Cut(V_k, V_{l'}-V_k)}(1-b_e)}{\prod_{e \in Cut(V_k, V_l-V_k)}(1-b_e)} \cdot \frac{q(l|V_k,\varphi',I,S)}{q(l'|V_k,\varphi,I,S)} \tag{12}$$

was defined for $\alpha(\varphi \rightarrow \varphi')$ (Eq. 11), the algorithmic process was ergodic, aperiodic, and had $p(W|I,S)$ as its stationary distribution, where $Cut(V_k, V_{l'} - V_k) = \{e_{ij}|v_i \in V_k, v_j \in (V_{l'} - V_k)\}$ was defined as the set of region edges between region group V_k and $V_{l'} - V_k$ (Fig. 3). As 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. 12 and the posterior probability $p(W|I,S)$ defined in Eq. 5.

4.2. Revised Band Probability and Proposed Move

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, and it determines how likely a pair of adjacent regions should be grouped together. Intuitively, b_{ij} should be large if the image information included in two regions is compatible, or if the shape of the merged region is more similar to the shape prior than either v_i or v_j. As a result, we define

$$\begin{aligned} b_{ij} &\propto p(I|v_i,v_j)\,p(S|v_i,v_j) \\ &\propto e^{-\mathcal{M}_I(I_{v_i},I_{v_j})} \cdot e^{\mathcal{M}_S(v_i,v_j,S)} + \epsilon \end{aligned} \tag{13}$$

where $\mathcal{M}_S(v_i,v_j,S)$

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

with

$$\eta = \min\{1, \ell_c(v_{i\cup j},S)/\max\{\ell_c(v_i,S),\ell_c(v_j,S)\}\}, \tag{14}$$

and $\mathcal{M}_I(I_{v_i},I_{v_j})$

$$\propto \sum_k \frac{(H_i(k)-m(k))^2}{m(k)}, m(k) = \frac{H_i(k)+H_j(k)}{2}. \tag{15}$$

As illustrated in Fig. 4, $v_{i\cup j}$ is a new region introduced by merging v_i and v_j; $\ell_c(v_{i\cup j},S)$ 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 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. For instance, while $M(v_{i\cup j},S)$ is smaller than $\max(M(v_i,S),M(v_j,S))$, η could be greater than 1 when $v_{i\cup j}$ was matched with the shape prior in a large scale.

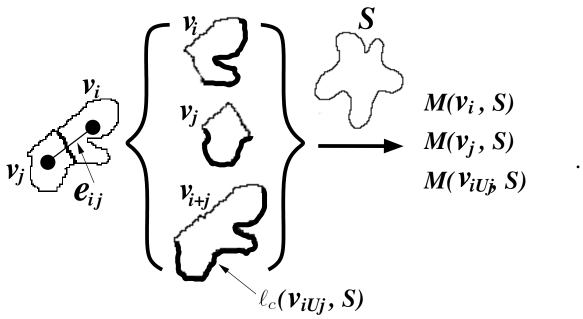


Figure 4: **Shape-Based Band Probability:** The longest partially matched boundaries of the image regions are shown in bold.

$\mathcal{M}_I(I_{v_i}, I_{v_j})$ measures the image compatibility between the region cells v_i and v_j. In the current implementation, we compute n-bin color histograms $H_i(n)$ and $H_j(n)$ for each of v_i and v_j, respectively. Many dissimilarity measurements can be applied to compute the distance between two histograms [14]. In the current implementation, $\mathcal{M}_I(I_{v_i}, I_{v_j}, I)$ is computed as a χ^2 statistics distance.

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

$$\{V_1, V_2, \ldots, V_l - V_k, \ldots, V_n, \varnothing\} \tag{16}$$

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, and $\varnothing$ is an empty set. The probability $q(l'|V_k, \varphi, I, S)$ defines how likely it is that V_k is merged with a region group $V_{l'}$ among all candidate groups, which is similarly defined as Eq. 13:

$$\begin{aligned} q(l'|V_k, \varphi, I, S) &= p(V_k \text{ is merged with } V_l') \\ &\propto p(I|V_k, V_{l'})\, p(S|V_k, V_{l'}) \end{aligned} \tag{17}$$

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

Because the band probability (Eq. 13) and the proposed move (Eq. 17) characterize the dominant properties modeled by the posterior probability (Eq. 5) well, they allow the proposed move $\varphi \rightarrow \varphi'$ to be accepted with a high probability such that the designed Markov chain will quickly converge to the expected solution.

4.3. Large Sampling Jump by Shape Registration

The shape similarity measurement computed in Section 3.1 could provide an additional correspondence constraint between the matched image region and the shape prior. A well-established correspondence constraint is very useful for performing the ideal shape registration in general, which may allow the simulation process to realize a large sampling jump. Given an observed image region V and the shape prior S, the sets of the correspondence points on their matched boundaries can be recorded as P_V and P_S, which respectively belong to the region boundary and the shape prior. A pair (t, θ) of translation and rotation parameters can be calculated by the least-squares method for registering P_S with P_V. Based on the obtained (t, θ), the prior shape S in the matched scale can be transformed onto the image and noted as S_T. A set of atomic regions inside this transformed shape prior S_T is recorded as:

$$\{v_i | v_i \in S_T, \text{for all } i\}, \tag{18}$$

where the operation $\in$ is simply implemented by judging if the centroid of v_i is inside S_T. A large region group V_Σ is then generated by merging all atomic regions in the above set (Eq. 18). When the size of V_Σ is comparable to the shape prior in the matched scale, we can further compute the relevant posterior probabilities (Eqs. 8 and 9) for V_Σ. V_Σ is recorded as the recognized object when the obtained posterior probability is larger than a given threshold, for instance 0.8, depending on what degree of occlusion is allowed for the objects to be retrieved in the current method.

However, the obtained segmentation state via a large sampling jump does not correspond to the states normally reached by the designed Markov chain. Such a sampling jump is not reversible in general and could cause the result be biased from its optimal solution. Therefore, in the current implementation, we embedded this operation within each proposed move and checked if any matched object could be recognized, while continuing with the Metropolis-Hasting sampling in its usual way.

5. System Implementation

The main steps of MosaicShape algorithm were implemented as follows. For an given input image I, a shape prior S, a learned image model θ_I and other segmentation parameters P, the system first applied a mean shift method [4] to generate the oversegmented images (Step 1), in which a bandwidth value and the minimum region size were chosen so that the number of generated oversegmented regions was kept moderately small. The boundary points of each region were identified and linked into a 2D closed contour. The curvature values of each contour were computed by Eq. 1, where a Fast Fourier Transformation (FFT) and inverse FFT were applied to speed up the convolution operations (Step 2). Based on these curvatures, the band probabilities (Eq. 13) between adjacent atomic regions were computed via the PartialShapeCorrelation algorithm (Step 3). A sampling move from the current state φ to the next φ' was implemented by a Metropolis-Hastings mechanism (Steps 4–7). During this process, the atomic regions lying within a previously well matched region group were chosen as v_k with a very small probability (Step 4.1). Within every proposed move, the recent region group and a large sampling jump were checked, and the matched object was recorded (Step 6). The evaluation of probabilities defined in Eqs. 5, 13 and 17 was implemented by a regularization framework, where a small weight (0.2) was associated with the relevant image probability, and a large weight (0.6) was assigned for the shape relevant probability. Moreover, to speed up the overall simulation process, for some operations requiring an expensive computation, such as curvature calculation, partial shape matching, or color histogram construction, these results were only computed once for a newly generated region group, and they were then stored into a sorted linked list (or hash table) indexed by the labels of the inner atomic regions of the current region group. Afterwards, when the same region group was visited, these results could be accessed efficiently by a search with a 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, respectively. Synthetic images were then created by randomly placing several mosaic "star" objects onto real images that included a lot of clutter in their backgrounds. Fig. 5 shows the segmentation results for detecting the star objects with a complete shape in different rotations and scales. Fig. 6 shows the results of detecting the objects with only partially matched shapes. The desired partial objects were retrieved by specifying an acceptable shape similarity threshold – in this experiment 0.7.

To test the current method on real images, an ellipse shape prior was first learned from a set of 12 leaves objects (Fig. 7). We then used some fallen leaves as the objects to be detected in the real images. 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, during the large sampling jump process, the registration parameters (t, θ) were first computed between the image regions and the learned mean shape. These parameters were then applied for registering each shape sample in the training set onto the image. Some experimental results for detecting the objects with complete or partially matched shapes are shown in Figs. 8 and 9.

The overall segmentation process took 30–90 s for synthetic and 150–250 s for real images. 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 be responsible for situations when the current method fails (Fig. 9)2(c).

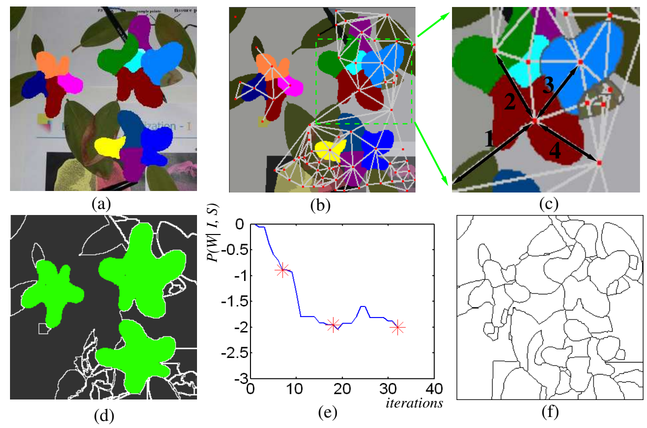


Figure 5: (a) Input image. (b) Oversegmented image with region adjacency graph. (c) Close-up view for some local region. The respective band probabilities for edges 1–4 are 0.24, 0.68, 0.67 and 0.39 (Eq. 14). (d) Segmentation result (green). (e) Posterior energy during the simulation. In red, sampling states at which the target objects were identified. (f) Segmentation result by DDMCMC method [18].

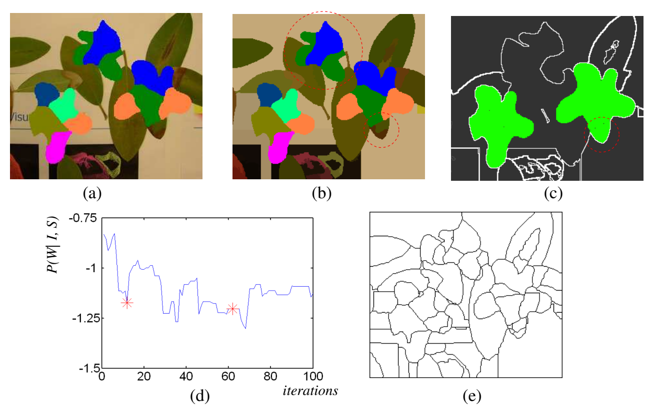


Figure 6: (a) Input image. (b) Oversegmented image. Object in large red circle has partial shape matching similarity 0.64. (c) Our result. Atomic regions in small red circle were included in a matched object due to a large sampling jump. (d) Posterior energy and two sampling states at which two objects with a partial shape similarity larger than 0.7 were found. (e) Segmentation result by DDMCMC method [18].

7. Discussion and Conclusion

The paper proposed a novel framework for object segmentation and recognition. The main contribution was to integrate the decomposed shape constraints into a bottom-up image segmentation process by a partial shape matching

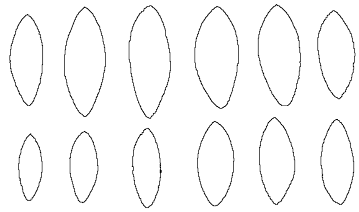


Figure 7: The shape samples of a leave object.

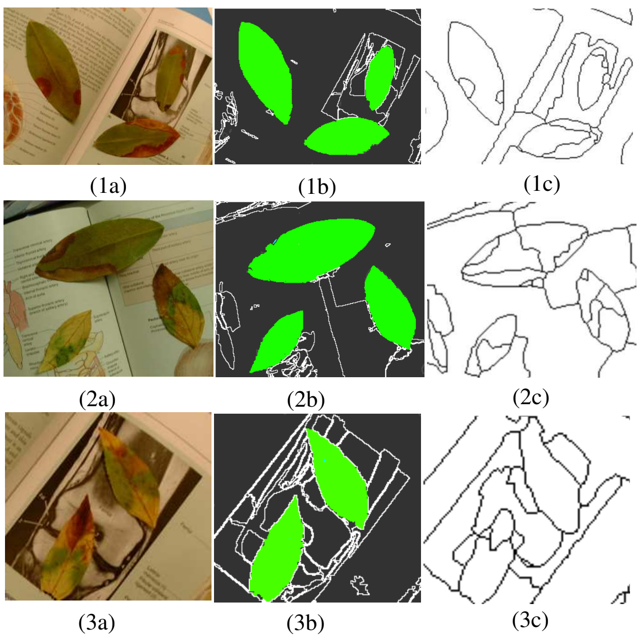


Figure 8: 1st column: Input images; 2nd column: Our results; 3rd column: Results by DDMCMC [18].

process. As a result, the segmentation and recognition of multiple occluded objects can be achieved simultaneously. The current method may be improved in the following aspects. First, a better shape similarity measurement might help produce more accurate results. For instance, a dynamic time warping method [12] may be applied to support partial shape matching for objects with distorted shapes. Second, the variability within the class of the shape prior may be modeled by principle component analysis, however, additional challenges may be encountered in applying partial shape matching. Finally, integrating comprehensive image models [18, 13] into the current system may help capture additional variabilities of the object appearance and thus obtain improved segmentation results.

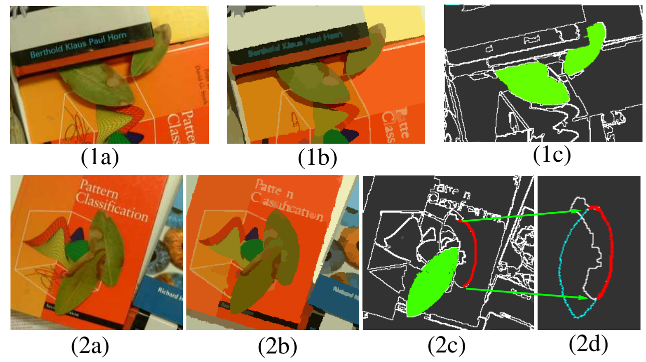


Figure 9: 1(a)-2(a): Input images; 1(b)-2(b): Oversegmented images; 1(c)-2(c): Our results. The matching ambiguity inherent in the partial shape match problem could directly affect the process of recognizing the object of interest. As shown in 2(d), the occluded leave with a partially matched boundary (marked in red) could lead into an unexpected shape registration result (marked in blue).

References

[1] A. Barbu and S. Zhu. Graph partition by Swendsen-Wang cuts. In *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 03)*, 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 *Proceedings of IEEE Workshop on Perceptual Organization in Computer Vision*, Washington DC, USA, 2004.

[3] Y. Chen, H. Tagare, and etc. Using prior shapes in geometric active contours in a variational framework. *Int J Comput Vis*, 50:315–328, 2002.

[4] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[5] A. K. Jain, Z. Yu, and S. Lakshmanan. Object matching using deformable templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(3):267–278, 1996.

[6] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int J Comput Vis*, 1:321–331, 1987.

[7] J. Kaufhold and A. Hoogs. Learning to segment images using region-based perceptual features. In *CVPR*, pages 954–961, Washington DC, USA, 2004.

[8] M. E. Leventon, W. E. L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, Volume I*, pages 316–323, Hilton Head, SC, USA, June 2000.

[9] S. Loncaric. A survey of shape analysis techniques. *Pattern Recognition*, 31(8):983–1001, 1988.

[10] R. Malladi, J. Sethian, and B. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:158–175, 1995.

[11] F. Mokhtarian. Silhouette-based isolated object recognition through curvature scale space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):539–544, 1995.

[12] E. Petrakis, A. Diplaros, and E. Millos. Matching and retrieval of distorted and occluded shapes using dynamic pro-

gramming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(11):1501–1516, 2002.

[13] X. Ren and J. Malik. Learning a classification model for segmentation. In *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 03)*, pages 10–17, Nice, France, Oct. 2003.

[14] Y. Rubner, C. Tomasi, and L. Guibas. The earth mover's distance as a metric for image retrieval. *Int J Comput Vis*, 40:99–121, 2000.

[15] S. Sclaroff and L. Liu. Deformable shape detection and description via model-based region grouping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(5):475–489, 2001.

[16] R. Swendsen and J. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Phys Rev Lett*, 58(2):86–88, 1987.

[17] Z. W. Tu, X. R. Chen, A. L. Yuille, and S. C. Zhu. Image parsing: unifying segmentation, detection, and recognition. In *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 03)*, pages 18–25, Nice, France, Oct. 2003.

[18] Z. W. Tu and S. C. Zhu. Image segmentation by data-driven Markov Chain Monte Carlo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(5):657–673, 2002.

[19] 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.

[20] G. Winkler. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*. Springer Verlag, 1995.

[21] A. Yuille, D. Cohen, and P. Hallinan. Feature extraction from faces using deformable templates. *Int J Comput Vis*, 8(2):99–111, 1992.