# A Bayesian Approach for Image Segmentation with Shape Priors [*]

Hang Chang[1,2], Qing Yang[1], and Bahram Parvin[2]
[1]Institute of Automation, Chinese Academy of Sciences
[2]Lawrence Berkeley National Laboratory
*hchang@lbl.gov*

## Abstract

*Color and texture have been widely used in image segmentation; however, their performance is often hindered by scene ambiguities, overlapping objects, or missing parts. In this paper, we propose an interactive image segmentation approach with shape prior models within a Bayesian framework. Interactive features, through mouse strokes, reduce ambiguities, and the incorporation of shape priors enhances quality of the segmentation where color and/or texture are not solely adequate. The novelties of our approach are in (i) formulating the segmentation problem in a well-defined Bayesian framework with multiple shape priors, (ii) efficiently estimating parameters of the Bayesian model, and (iii) multi-object segmentation through user-specified priors. We demonstrate the effectiveness of our method on a set of natural and synthetic images.*

## 1. Introduction

Segmentation has been a fundamental topic in computer vision and image understanding. Since a full review of segmentation methods is beyond the scope of this paper, we briefly address two important computational modules that relate to our paper. These are level set and graph cut methods. The level set method of Osher and Sethian [16] has been widely applied and extended for low-level image analysis. The main advantage of the level set formulation is to define the curve evolution in a Eucledean framework, which improves the numerical stability and ensures the topological change. Extensions and evolutions of the level set method have been to (i) express the curve evolution in terms of an optimization framework, which is known as the "Geodesic active contour" model [4]; (ii) handle regions with weak boundaries via region-based method of Chan and Vese [6]; and (iii) incorporate the shape prior model [8, 5] with invariance to scaling, translation, and rotation.

The graph cut mehtod is a combinatorial optimization method, which arrives at a fixed point through global optimization for a certain class of energy functions. Recently, based on graph cut algorithm [3], successful interactive image segmentation strategies [17, 15, 1] have been proposed, where user interaction is reduced to a few mouse-strokes indicating foreground and background features. To incorporate shape prior into the graph cut approach, Freedman and Zhang [10] integrated information about the level set function of a template with the traditional graph edge-weights so that the edges in the graph could convey information about both the image and the shape prior.

Although the ultimate objective of the community is in automated segmentation and interpretation, which can be regarded as quasi object recognition, we have opted for an interactive segmentation procedure as an interim step. As a result of reducing inherent ambiguities and complexities associated with automated segmentation, we can (i) test and evaluate building blocks in a controlled fashion, and (ii) build robust systems. In this paper, we propose a new algorithm for interactive image segmentation with shape priors within a Bayesian framework. The framework allows the use of multiple shape priors to obtain the best fit; affine transformation of the model; and multi-objects segmentation within the same image.

The rest of our paper is organized as follows: In Section 2, we briefly introduce Freedman and Zhang's paper, and the basic graph cut method. In Section 3, we introduce an energy function with the shape prior model, which is important for the probabilistic model definition. In Section 4, we outline the details of the Bayesian optimization approach. Section 5 provides experimental results and comparisons. Section 6 concludes the paper.

## 2. Related Work

In this section, we review the principal of the graph cut method for image segmentation and Freedman and Zhang's work [10]. The graph cut was first introduced by Greig [11] as a combinatorial optimization method in the context of the max-flow/min-cut of a graphical network for a class of

energy functionals. Following a number of recent papers [3, 19, 14, 18], the graph cut algorithm has emerged as an increasingly powerful tool for energy minimization in low-level vision problems. For the segmentation problems, it works as a binary labeling approach based on the graph $G = \langle \bar{V}, \bar{E} \rangle$ constructed from the image, where $\bar{V}$ is the set of all nodes, and $\bar{E}$ is the set of all arcs connecting adjacent nodes. Usually, the nodes are pixels in the image, and arcs are adjacency relationships with four or eight connections between neighboring pixels. In addition, there are some special nodes called terminals in the graph structure. In the context of segmentation, terminals correspond to the set of labels that can be assigned to pixels. We will concentrate on the case of a graph with two terminals, which are usually called the source ($\mathcal{S}$) and the sink ($\mathcal{T}$). The labeling approach is to assign an unique label $x_p$ (0 for background, and 1 for foreground ) for each node $p \in \bar{V}$, and thus the image cutout is performed by minimizing the Gibbs energy $E(X)$ [12]:

$$E = \sum_{p \in \bar{V}} E_1(x_p) + \sum_{(p,q) \in \bar{E}} E_2(x_p, x_q) \qquad (1)$$

Where $E_1(x_p)$ is the likelihood energy, encoding the fitness cost for assigning $x_p$ to $p$, and $E_2(x_p, x_q)$ is the prior energy, denoting the cost when the labels of adjacent nodes $p$ and $q$ are $x_p$ and $x_q$, respectively. There are polynomial algorithms for the optimization on directed weighted graphs with two terminals. These algorithms could be classified into two groups: Goldberg-Tarjan "push-relabel" methods[13] and Ford-Fulkerson "augmenting paths" [9]. The details of the methods could be found in [7].

Based on the graph cut formulation, Freedman and Zhang [10] integrated the information about a level set function of a template with the traditional graph's edge-weights to allow the edges of the graph to convey information about both the image and the shape prior. They used the following objective energy function:

$$
\begin{aligned}
E \;=\;& \sum_{p \in P} (1 - \lambda) \mu R_p(A(p)) \qquad (2) \\
& + \sum_{(p,q) \in \mathbb{N}: A(p) \neq A(q)} \left[ (1 - \lambda) B_{p,q} + \lambda \bar{\phi}(\frac{p+q}{2}) \right]
\end{aligned}
$$

where $p$ and $q$ are pixels in the image. $P$ is the set of all pixels. $\mathbb{N}$ is the neighborhood system. $A$ is the label of the segmented image: $A(p) = 1$, if $p \in Foreground$ and $A(p) = 0$, if $p \in Background$. $\lambda(0 \le \lambda \le 1)$ and $\mu(\mu \ge 0)$ are the weights for different parts. And $\bar{\phi}$ is the unsigned distance function of the shape prior model that is derived from its contour. Additional details of the energy terms can be found in [10]. To tackle the transformation of the template, Freedman and Zhang mapped the center of the user's foreground strokes to the centroid of the object,

and used a variant of the Procrustes Method [20] to further estimate the rotation parameter. Following translation and rotation of the template, they computed a Gaussian pyramid of the image, and performed graph cut segmentation on each level of the pyramid, whereby the cut with minimum normalized energy leads to the desired segmentation.

## 3. Incorporation of the Shape Model

In this section, we first introduce an energy function with the shape prior for subsequent use in the probabilistic framework. This is followed with the data and smoothness constraints.

### 3.1. Shape Prior Model

Given a shape prior, which is delineated by a closed curve $c$ (or multiple closed curves, if there are holes inside), we create a labeling function, $L$, shown in Figure 1 as:

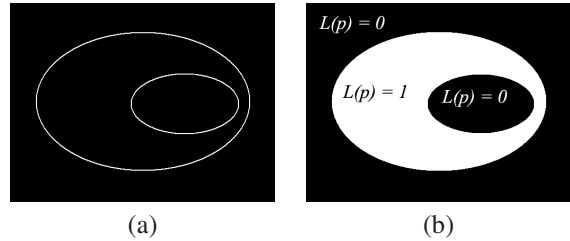$$L(p) = \begin{cases} 1, & \text{if } p \text{ is within the shape region} \\ 0, & \text{Otherwise} \end{cases}$$



Figure 1. An example of a labeling function from a given shape.

We define the energy function with the shape prior model as

$$
\begin{aligned}
E(t, sp, A) \;=\;& \sum_{p \in P} (1 - \lambda) E(A(p)) \\
& + \sum_{(p,q) \in \mathbb{N}} (1 - \lambda) \mu \cdot w_e(p, q) \cdot \delta_{A(p), A(q)} \\
& + \sum_{p \in P} \lambda \delta_{A(p), L_{t,sp}(p)} \qquad (3)
\end{aligned}
$$

where $A, \lambda, \mu,$ and $\mathbb{N}$ have the same meanings as in (2); $sp$ refers to the shape prior; $t$ is a configuration of the affine transformation matrix; $L_{t,sp}$ is the transformed labeling function; and

$$\delta_{X,Y} = \begin{cases} 0, & \text{if } X = Y \\ 1, & \text{if } X \neq Y \end{cases}$$

The first part of Equation (3) is the data fitness term; the second component is the smoothness term; and the third part is the shape fitness term, which is minimum when $A =$

$L_{t,sp}$. The third term matches the segmentation result ($A$) and the labeling function ($L_{t,sp}$), and is controlled by $\lambda$.

The affine transformation matrix, $T$, could be written as

$$T = \begin{pmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ 0 & 0 & 1 \end{pmatrix} \tag{4}$$

$$\begin{aligned} a_{11} &= S_x \cos\theta - S_x SH_x \sin\theta \\ a_{12} &= S_x \sin\theta + S_x SH_x \cos\theta \\ a_{21} &= S_y SH_y \cos\theta - S_y \sin\theta \\ a_{22} &= S_y SH_y \sin\theta + S_y \cos\theta \end{aligned}$$

where $S_x$ and $S_y$ are the scaling factors (in the $x$ and $y$ dimensions); $SH_x$ and $SH_y$ are the shearing factors; $b_1$ and $b_2$ are the translation parameters; and $\theta$ is the rotation parameter. Hence, $L_{t,sp}$ is related to $L$ (labeling function of $sp$) by

$$\begin{aligned} L_{t,sp}(x,y) &= L(x^\star, y^\star), \tag{5} \\ x^\star &= \frac{a_{22}(x - b_1) - a_{12}(y - b_2)}{a_{11}a_{22} - a_{12}a_{21}}, \\ y^\star &= \frac{a_{21}(x - b_1) - a_{11}(y - b_2)}{a_{12}a_{21} - a_{11}a_{22}} \end{aligned}$$

In contrast to Freedman and Zhang's formulation [10], we do not compute the distance map for $c$, and the shape fitness term corresponds only to the amount of mismatches with the labeling function, regardless of their location. As a result, our method avoids the "translation" and "scale" ambiguities caused via Freedman and Zhang's formulation:

1. *Translation ambiguity:* Two segments should have the same shape fitness energy if they are identical and within the region of the prior shape. As shown in Figure 2(a), the shape fitness energies for both segments are the same in our approach, but they are different in other formulations [10].

2. *Scale ambiguity:* If a segment has the same shape and centroid as that of the shape prior, but with different scale, then we represent the scale difference as $\triangle_S(\geq 0)$. The shape fitness energy function should monotonically increase with respect to $\triangle_S$. Assume, w.l.g, the shape prior and the candidate segment are circular with radius $r$ and $\beta r (0 \leq \beta \leq 1)$, respectively. In the continuous case, the shape fitness energy in [10] is $2\pi r^2 \cdot \beta(1 - \beta)$; here, we drop the constant $\lambda$ for clarity. Thus, two segments with $\beta_1 = (\frac{1}{2} + \alpha)$ and $\beta_2 = (\frac{1}{2} - \alpha)$ $(0 < \alpha \leq \frac{1}{2})$ have the same shape fitness energy, as shown in Figure 2(b), which causes the scale ambiguity.
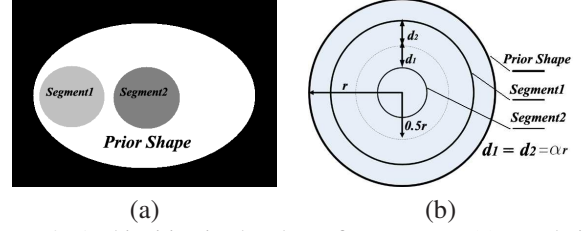


Figure 2. Ambiguities in the shape fitness term: (a) translation ambiguity; and (b) scale ambiguity.

## 3.2. Data and Smoothness Model

In the case of the data fitness term, we establish two Gaussian Mixture Models ($GMM$) based on user strokes to delineate the foreground and background features, which could be color ($RGB$) or texture (extracted by Gabor filter banks). $E(A(p))$ is defined as,

$$E(A(p)) = \begin{cases} \frac{GMM_B(f_p)}{GMM_F(f_p) + GMM_B(f_p)}, & \text{if } A(p) = 1 \\ \frac{GMM_F(f_p)}{GMM_F(f_p) + GMM_B(f_p)}, & \text{if } A(p) = 0 \end{cases}$$

in which $GMM_F(f_p)$ and $GMM_B(f_p)$ are the probabilities of feature $f_p$ from the foreground and background models, respectively. The smoothness term is defined as,

$$w_e(p,q) \propto exp\left(\frac{-dist(f_p, f_q)}{2\sigma^2}\right)$$

where $dist(f_p, f_q)$ is the Euclidean distance between feature vectors of $f_p$ and $f_q$. $\sigma$ is kept as a constant($\sigma = 3$) for all experiments except for the synthetic texture image, in which $\sigma$ is set to be 6.

## 4. Optimization of the Bayesian Model

We begin our approach with user interactions, the details of which are covered in Section 4.3. Next, we formulate the problem of interactive image segmentation with shape priors in a well-defined Bayesian framework, which is solved via the maximum a posterior ($MAP$) technique. $MAP$ estimates $(t, sp, A)$ given the image $I$ and the set of shape priors, $SP$, in which $sp \in SP$. It is a maximization over a probability distribution $P$ and can be expressed via Baye's rule as:

$$\begin{aligned} & \arg\max_{t,sp,A} P(t, sp, A | SP, I) \tag{6} \\ =\ & \arg\max_{t,sp,A} P(SP, I | t, sp, A) P(A | t, sp) \cdot \\ & P(t, sp) / P(SP, I) \\ =\ & \arg\max_{t,sp,A} P(SP, I | t, sp, A) P(A | t, sp) P(t, sp) \end{aligned}$$

where $P(SP, I)$ has been dropped, because it is a constant with respect to the optimization. We now outline every component of the probabilistic modeling and optimization step.

## 4.1. Probabilistic Models

1. The $P(t, sp)$ Model describes the probability of shape prior, $sp$, with transformation parameters, $t$. For each $(t, sp)$, we construct a partition to exactly match the transformed shape prior: $A = L_{t,sp}$. The energy of this partition could be calculated via (3), and the model is defined as:

$$P(t, sp) = \frac{E(t, sp, A = L_{t,sp})^{-1}}{\sum_{sp' \in SP} \int_T E(t', sp', A = L_{t',sp'})^{-1} dt'} \quad (7)$$

in which $T$ is all combinations of transformation parameters. Since segmentation could be formulated as an energy minimization approach, the above definition indicates that the combination, $(t, sp)$, with smaller energy of the pre-segmentation, is more likely to lead to the desired shape prior and its transformation.

2. The $P(A|t, sp)$ Model describes the probability of segmentation $A$, under the condition of shape prior, $sp$, and its transformation parameters, $t$, and is defined as,

$$P(A|t, sp) = \frac{E(t, sp, A)^{-1}}{\sum_{A' \in \mathbb{A}} E(t, sp, A')^{-1}} \quad (8)$$

in which $\mathbb{A}$ is the set of all possible partitions. For two-region segmentation, the number of all possible partitions is $2^K$, in which $K$ is the total number of pixels in the image. This definition means that, with fixed $(t, sp)$, the partition with smaller energy is more likely to be the desired segmentation.

3. The $P(SP, I|t, sp, A)$ Model takes advantage of the independence between the shape prior and image content. Thus, it is modeled as,

$$P(SP, I|t, sp, A) = P(SP|t, sp, A)P(I|t, sp, A)$$

$$P(SP|t, sp, A) \propto \exp\left(\frac{-\| L_{t,sp} - A \|}{\| L_{t,sp} \|}\right)$$

$$P(I|t, sp, A) \propto \prod_{p \in P} E\big(1 - A(p)\big)$$

where $P(SP|t, sp, A)$ models the fitness between the segmentation and the transformed shape prior, $P(I|t, sp, A)$ represents the data fitness under the current label configuration, and

$$\| L_{t,sp} \| = \sum_{p \in P} L_{t,sp}(p)$$

$$\| L_{t,sp} - A \| = \sum_{p \in P} \delta_{A(p), L_{t,sp}(p)}$$

## 4.2. Optimization Procedure

Equation (6) is optimized iteratively:

1. Initialization: Construct two $GMM$s based on user strokes delineating foreground and background features, respectively. $Iter = 0$.

2. **IF** ($Iter = 0$), estimate the best fit $(t, sp)$, which leads to $\max_{t,sp} P(t, sp)$.
   **ELSE**, refine transformation parameter, $t$, based on estimated $(t, sp)$ and $A$ to maximize $P(SP|t, sp, A)$.

3. Estimate the partition $A$ that leads to $\max_A P(A|t, sp)$ with estimated $(t, sp)$. $Iter = Iter + 1$.

4. Repeat steps 2 and 3, until (6) remains unchanged.

Details of each step are described as follows:

### 4.2.1 Estimation for $(t, sp)$

By definition, the maximum $P(t, sp)$ will be based on the minimum $E(t, sp, A = L_{t,sp})$, and the shape fitness term in (3) has no impact. Therefore, the problem could be refined as the estimation of $(t, sp)$, which leads to $\min_{t,sp} E(t, sp, A = L_{t,sp})$. Here, we rewrite the objective energy function as,

$$\begin{aligned} E(t, sp, A) &= \sum_{(x,y)\in\Omega} E(A(x, y) = 1) \cdot L_{t,sp}(x, y) \quad (9) \\ &+ \sum_{(x,y)\in\Omega} E(A(x, y) = 0) \cdot (1 - L_{t,sp}(x, y)) \\ &+ \mu \sum_{(x,y)\in\Omega} w_e \cdot |\nabla L_{t,sp}(x, y)| \end{aligned}$$

where, $A = L_{t,sp}$; $\Omega$ is the image domain; and the constant $(1 - \lambda)$ has been dropped for clarity. Our analysis indicates that the smoothness energy in (3) has little impact on the estimation of $(t, sp)$, because once shape information is necessary for improving robustness, it means that the edge information may not be reliable due to overlapping objects or ambiguities in feature-based representation. Hence, by setting $\mu = 0$, we can simplify the estimation problem. The objective energy is minimized by fixing $sp$, and applying the gradient descent with respect to the transformation parameters:

$$\begin{aligned} \frac{\partial S_x}{\partial t} &= \sum_{\Omega} K \cdot L_{x^\star} \frac{(SH_y \sin\theta + \cos\theta)(x - b_1)}{S_x^2(1 - SH_x SH_y)} \\ &- \sum_{\Omega} K \cdot L_{y^\star} \frac{(SH_y \cos\theta - \sin\theta)(x - b_1)}{S_x^2(1 - SH_x SH_y)} \end{aligned}$$

$$(10)$$

$$\frac{\partial S_y}{\partial t} = -\sum_\Omega K \cdot L_{x^\star} \frac{(SH_x \cos\theta + \sin\theta)(y - b_2)}{S_y^2(1 - SH_x SH_y)}$$

$$- \sum_\Omega K \cdot L_{y^\star} \frac{(SH_x \sin\theta - \cos\theta)(y - b_2)}{S_y^2(1 - SH_x SH_y)}$$

$$\frac{\partial SH_x}{\partial t} = \sum_\Omega K \cdot L_{x^\star} \frac{(SH_y \sin\theta + \cos\theta)(y - b_2)}{S_y(1 - SH_x SH_y)^2}$$

$$- \sum_\Omega K \cdot L_{x^\star} \frac{SH_y(SH_y \sin\theta + \cos\theta)(x - b_1)}{S_x(1 - SH_x SH_y)^2}$$

$$+ \sum_\Omega K \cdot L_{y^\star} \frac{SH_y(SH_y \cos\theta - \sin\theta)(x - b_1)}{S_x(1 - SH_x SH_y)^2}$$

$$- \sum_\Omega K \cdot L_{y^\star} \frac{(SH_y \cos\theta - \sin\theta)(y - b_2)}{S_y(1 - SH_x SH_y)^2}$$

$$\frac{\partial SH_y}{\partial t} = -\sum_\Omega K \cdot L_{x^\star} \frac{(SH_x \cos\theta + \sin\theta)(x - b_1)}{S_x(1 - SH_x SH_y)^2}$$

$$+ \sum_\Omega K \cdot L_{x^\star} \frac{SH_x(SH_x \cos\theta + \sin\theta)(y - b_2)}{S_y(1 - SH_x SH_y)^2}$$

$$- \sum_\Omega K \cdot L_{y^\star} \frac{(SH_x \sin\theta - \cos\theta)(x - b_1)}{S_x(1 - SH_x SH_y)^2}$$

$$+ \sum_\Omega K \cdot L_{y^\star} \frac{SH_x(SH_x \sin\theta - \cos\theta)(y - b_2)}{S_y(1 - SH_x SH_y)^2}$$

$$\frac{\partial \theta}{\partial t} = -\sum_\Omega K \cdot L_{x^\star} \frac{(SH_y \cos\theta + \sin\theta)(x - b_1)}{S_x(1 - SH_x SH_y)}$$

$$- \sum_\Omega K \cdot L_{x^\star} \frac{(SH_x \sin\theta - \cos\theta)(y - b_2)}{S_y(1 - SH_x SH_y)}$$

$$- \sum_\Omega K \cdot L_{y^\star} \frac{(SH_y \sin\theta + \cos\theta)(x - b_1)}{S_x(1 - SH_x SH_y)}$$

$$+ \sum_\Omega K \cdot L_{y^\star} \frac{(SH_x \cos\theta + \sin\theta)}{S_y(1 - SH_x SH_y)}$$

$$\frac{\partial b_1}{\partial t} = \sum_\Omega K \cdot L_{x^\star} \frac{SH_y \sin\theta + \cos\theta}{S_x(1 - SH_x SH_y)}$$

$$- \sum_\Omega K \cdot L_{y^\star} \frac{SH_y \cos\theta - \sin\theta}{S_x(1 - SH_x SH_y)}$$

$$\frac{\partial b_2}{\partial t} = -\sum_\Omega K \cdot L_{x^\star} \frac{SH_x \cos\theta + \sin\theta}{S_y(1 - SH_x SH_y)}$$

$$- \sum_\Omega K \cdot L_{y^\star} \frac{SH_x \sin\theta - \cos\theta}{S_y(1 - SH_x SH_y)}$$

where $L$ is the fixed labeling function representing the shape prior, $sp$, and $L_{t,sp}$ is related to $L$ via Equation 5, and

$$L_x = \frac{\partial L}{\partial x}, L_y = \frac{\partial L}{\partial y},$$

$$K = E(A(x,y) = 1) - E(A(x,y) = 0)$$

$t$ is initialized by: $S_x^0 = S_y^0 = 1$, $SH_x^0 = SH_y^0 = 0$, $b_1^0 = C_x, b_2^0 = C_y$, and $\theta^0 = 0$, where $(C_x, C_y)$ is the centroid of the user-specified rectangular region; in cases of single-object segmentation per image without a bounding box, $(C_x, C_y)$ is taken as the centroid of foreground strokes. For each $sp \in SP$, we perform the gradient descent approach to get a minimum energy and the corresponding transformation parameters, $t$. The $(t, sp)$, which leads to the minimum among the minimum energies, is chosen as the estimated $(t, sp)$.

#### 4.2.2 Estimation for $A$

Similar to the estimation of $(t, sp)$, maximizing $P(A|t, sp)$ is equivalent to minimizing $E(t, sp, A)$. In other words, the problem can be converted to estimating $A$, which leads to $\min_A E(t, sp, A)$ with fixed $(t, sp)$. We have opted to use the graph cut algorithm [2] for its efficiency. With estimated $(t, sp)$, we construct a classical two-terminal graph with weights, shown in Table 1, and perform the graph cut algorithm to get the optimal partition, $A$.

| Edge | Weight | For |
|------|--------|-----|
| | $\infty$ | $p \in F$ |
| $p \to \mathcal{S}$ | $0$ | $p \in B$ |
| | $(1 - \lambda)E(A_p = 0) + \lambda L'_p$ | $Otherwise$ |
| | $0$ | $p \in F$ |
| $p \to \mathcal{T}$ | $\infty$ | $p \in B$ |
| | $(1 - \lambda)E(A_p = 1) + \lambda(1 - L'_p)$ | $Otherwise$ |
| $w_{\{p,q\}}$ | $(1 - \lambda)\mu \cdot w_e(p, q)$ | $\{p, q\} \in \mathbb{N}$ |

Table 1. Edge weights for the graph construction. $F$ and $B$ are user-specified foreground and background samples, respectively. $A_p = A(p), L'_p = L_{t,sp}(p)$.

#### 4.2.3 Refinement for $t$ and $A$

After the first iteration, $t$ is refined to further maximize $P(SP|t, sp, A)$ via the gradient descent method, which is not listed in this paper, due to space limitation. With the refined $t$, the optimal partition, $A$, is re-computed, as described in Section 4.2.2. Notice that we did not use $P(I|t, sp, A)$ to refine $A$, the rationale being that, without shape information, this data fitness term could be misleading for the segmentation; however, it plays an important role for overall optimization by compromising between the shape fitness and data fitness. The iterative procedure stops when (6) is optimized.

### 4.3. Specification of User Interaction

Two types of user interactions are supported with different functionalities:

1. Boxed Regions: *Unnecessary for single-object segmentation per image*. The desired object is strictly within the rectangular region.

2. Stroke: *Necessary*. Pixels marked by "red" strokes and "blue" strokes are labeled as foreground and background features, respectively.

The roles of *"Boxed Region"* are to (i) initialize the centroid of the object, $(b_1^0, b_2^0)$, as the centroid of the rectangle, $(C_x, C_y)$; and (ii) enable the multi-object segmentation within one image by converting the problem into several sub segmentation problems within each rectangular region.

*Note: The strokes within one specific rectangular region are for that specific sub segmentation only. The strokes outside all rectangular regions are for all sub segmentations.*

## 5. Experimental Results

We have applied our approach to both natural and synthetic images incorporating complex color and texture features. In our experiments, we used an 8-neighborhood system for graph construction, and represented color in the $RGB$ space. Texture features were extracted by the Gabor filter banks, in which $\sigma \in \{2, 3\}$ and $\theta \in \{0, \frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{3}, \frac{5\pi}{6}\}$, and $\lambda$, $\mu$ for each experiment are listed in the caption of the corresponding figure. For multi-object segmentation, we constructed a graph for each rectangular region to compute the sub segmentation separately. All the shape priors were drawn by hand.

Figure 3 shows the segmentation of four leaves (three of them are maple leaves) in one image, where Figure 3(b) gives the shape priors used for segmentation, Figures 3(e)-(h) are the auto-selected priors with estimated poses by our approach, and Figures 3(i)-(l) indicate the segmentation results. For comparison, we set $\lambda = 0$ when computing $E(t, sp, A)$ to drop the shape prior in the estimation process, which led to the traditional graph cut method. Figure 3(d) shows the best segmentation result produced by traditional graph cut with the same foreground and background seeds, which is erroneous due to the ambiguities in the foreground and background features, whereas, this problem is well solved by our approach. Figures 4-6 show three more examples with seven segmented objects. Each of them has at least one challenging problem, such as feature ambiguity, or missing or overlapping parts, which further demonstrates the effectiveness of our approach.

Freedman and Zhang's approach has additional requirements of the strokes to indicate the center and pose of the object, which leads to increased sensitivity to initialization. Thus, for comparison, we emulated Freedman and Zhang's strokes to meet these requirements. Figure 7 shows an improved performance of our approach over [10], which is mainly due to our extended capture range for transforma-

tion, and iterative optimization strategy. Generally, we consider the work in [10] to be a specific, simplified case within our framework.

One might question how accurate the shape prior should be to ensure a satisfactory result. In our approach, the shape prior should express the main structure of the object within an affine transformation. For example, the shape prior for the "plate" in Figure 4 is an ellipse, and the segmentation of the letters in Figure 6 was based on priors with different fonts and affine transformation. However, there are cases that more accurate shape information is required to solve the ambiguity, such as the claws of the "cabrite" in Figure 5. With respect to the optimization efficiency, in all our cases, the method converged in less than 4 iterations.

## 6. Conclusion and Future Work

We have developed an interactive image segmentation approach with the shape prior models in a well-defined Bayesian framework. The method allows the incorporation of multiple shape priors, from which the best shape prior and its transformation parameters are estimated efficiently, and enables multi-object segmentation in one image. Our experimental results indicate that our method leads to satisfactory results in complex cases, where feature ambiguity, overlapping objects, and missing parts can exist. Future work will focus on the incorporation of (i) more complex shape deformation, and (ii) automatic shape detection.
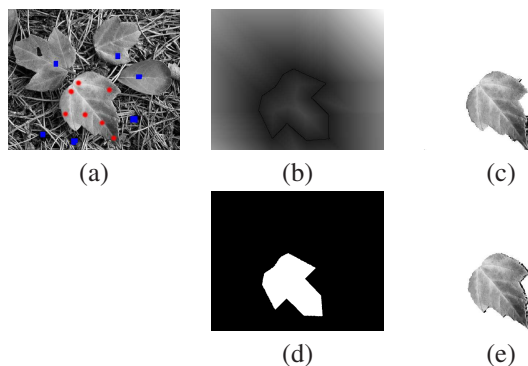


Figure 7. Comparison between our approach and the method in [10] based on the same user strokes and shape prior.(a) User strokes; (b) distance map estimated by the method in [10]; (c) segmentation result via the method in [10]; (d) $L_{t,sp}$ estimated by our approach; and (e) segmentation results via our approach.

## References

[1] Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In *Proc. of IEEE ICCV*, pages 105–112, 2001. 1
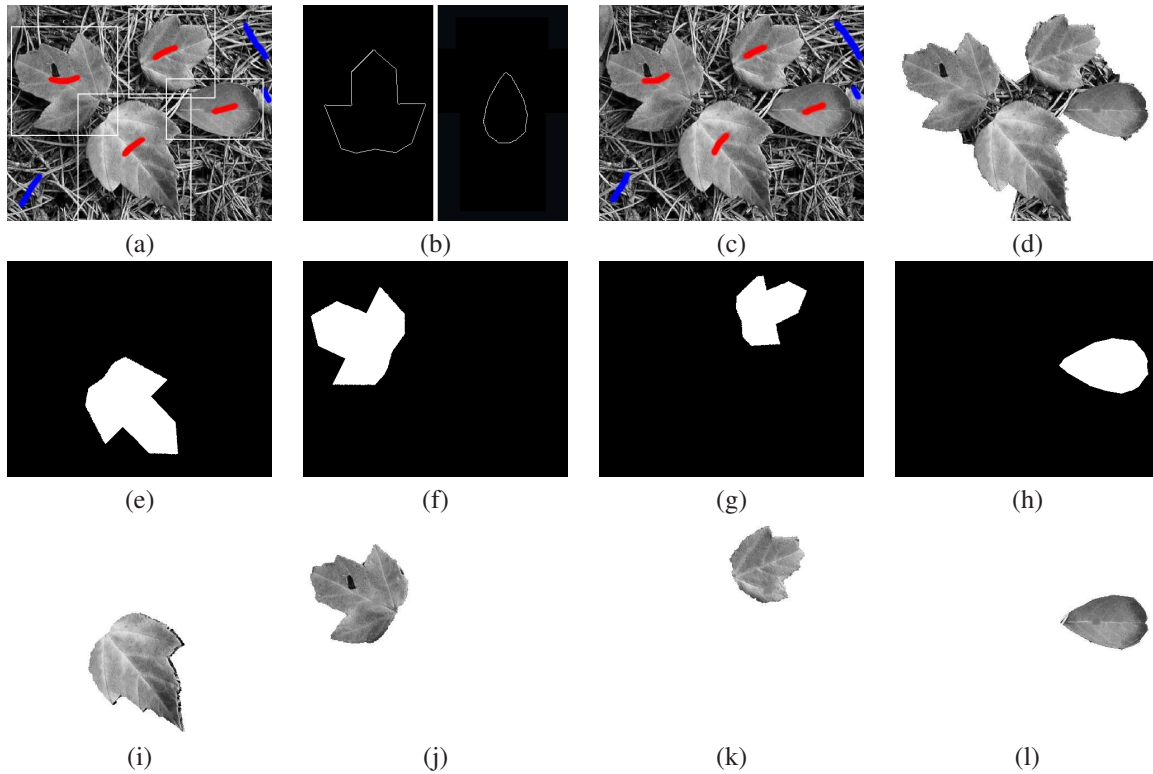
Figure 3. Segmentation for maple leaves with $\lambda = 0.2$, $\mu = 100$: (a) User input through mouse strokes and bounding box; (b) shape priors; (c) user interaction for graph cut; (d) segmentation result through graph cut; (e)-(h) estimated $L_{t,sp}$ for the four leaves; (i)-(l) results from the proposed method.
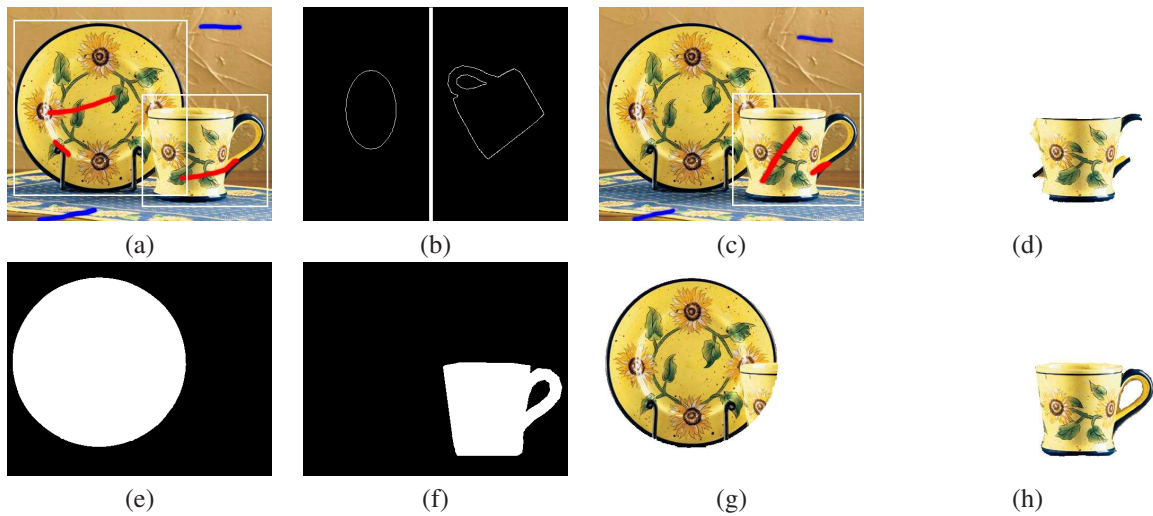


Figure 4. Segmentation for the mug and plate with $\lambda = 0.5$, $\mu = 100$: (a) User input through strokes and bounding box; (b) shape priors; (c) user input for graph cut; (d) segmentation result from graph cut; (e)(f) estimated $L_{t,sp}$; and (g)(h) results from the proposed method.

[2] Y. Boykov and V. Kolmogorov. An experimental comparision of min-cut/max-flow algorithms for energy minimization in vision. *In IEEE Transaction on PAMI*, 26(9):1124–1137, 2004. 5

[3] Y. Boykov, O. Veksler, and R. Zabih. Fast Approximate Energy Minimization via Graph Cuts. *In IEEE Transaction on PAMI*, 23(11):1222–1239, 2001. 1, 2

[4] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 694–699, 1995. 1
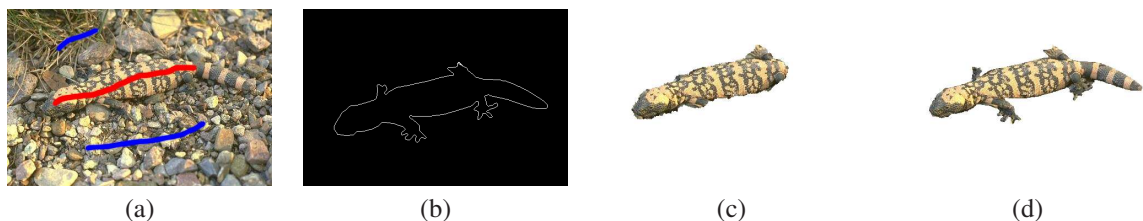
Figure 5. Segmentation for the cabrite image with $\lambda = 0.5$ and $\mu = 100$: (a) user input for both our method and the graph cut approach; (b) shape prior; (c) segmentation result via graph cut; and (d) segmentation result via with our approach.
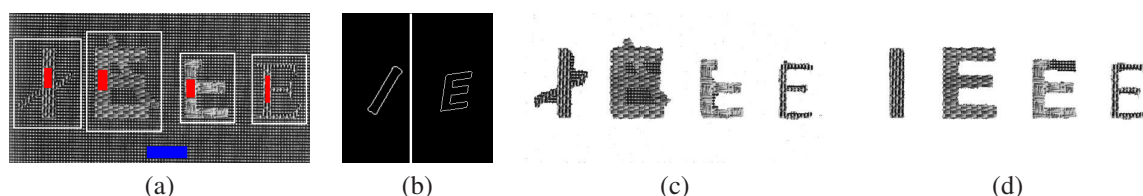


Figure 6. Segmentation for synthetic textured letters: The first three letters are in "Arial," the last one is in "Times New Roman", and the prior shapes are in "Arial" with $\lambda = 0.2$ and $\mu = 100$. (a) User input for both our method and the graph cut approach; (b) shape priors; (c) segmentation results via graph cut; and (d) segmentation results via our approach.

[5] T. Chan and W. Zhu. Level set based shape prior segmentation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 1164–1170, 2005. 1

[6] T.F. Chan and L.A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10:266–277, 2001. 1

[7] W.J. Cook, W.H. Cunningham, W.R. Pulleyblank, and A.Schrijver. *Combinatorial Optimization*. John Wiley & Sons, 1998. 2

[8] D. Cremers, N. Sochen, and C. Schnorr. Towards recognition-based variational segmentation using shape priors and dynamic labeling. In *Int. Conf. on Scale Space Theories in Computer Vision*, pages 388–400, 2003. 1

[9] L. Ford and D. Fullkerson. *Flows in Networks*. Princeton University Press, 1962. 2

[10] D. Freeman and T. Zhang. Interactive graph cut based segmentation with shape priors. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 755–762, 2005. 1, 2, 3, 6

[11] D. Geeig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *In Jorunal of the Royal Statistical Society*, 51(2):271–279, 1989. 1

[12] S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *In IEEE Transaction on PAMI*, 6(6):721–741, 1984. 2

[13] A. V. Goldberg and R. E. Tarjan. A new approach to maximum-flow problem. *In Journal of the Association for Computing Machinery*, 35(4):921–940, 1988. 2

[14] H. Ishikawa. Exact optimization for Markov Random Fields with convex priors. *In IEEE Transaction on PAMI*, 25(10):1333–1336, 2003. 2

[15] Y. Li, J. Sun, C.K. Tang, and H.Y. Shum. Lazy snapping. In *Proc. of ACM SIGGRAPH*, pages 303–308, 2004. 1

[16] S. Osher and J. Sethian. Front propagation with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1988. 1

[17] C. Rother, V. Kolmogorov, and A. Blake. Grabcut - Interactive foreground extraction using iterated graph cuts. In *Proc. of ACM SIGGRAPH*, pages 309–314, 2004. 1

[18] S. Roy. Stereo without epipolar lines: A maximum-flow formulation. *In International Journal of Computer Vision*, 34:147–162, 1999. 2

[19] V.Kolmogorov and R.Zabih. What energy functions can be minimized via graph cuts. *In IEEE Transaction on PAMI*, 26(2):147–159, 2004. 2

[20] T. Zhang and D. Freeman. Tracking objects using density matching and shape priors. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1056–1062, 2003. 2