Preferential Image Segmentation Using Trees of Shapes

Yongsheng Pan, *Member, IEEE*, J. Douglas Birdwell, *Fellow, IEEE*, and Seddik M. Djouadi, *Member, IEEE*

***Abstract*—A novel preferential image segmentation method is proposed that performs image segmentation and object recognition using mathematical morphologies. The method preferentially segments objects that have intensities and boundaries similar to those of objects in a database of prior images. A tree of shapes is utilized to represent the content distributions in images, and curve matching is applied to compare the boundaries. The algorithm is invariant to contrast change and similarity transformations of translation, rotation and scale. A performance evaluation of the proposed method using a large image dataset is provided. Experimental results show that the proposed approach is promising for applications such as object segmentation and video tracking with cluttered backgrounds.**

***Index Terms*—Curve matching, preferential image segmentation, tree of shapes.**

I. Introduction

IMAGE segmentation algorithms are designed to segment an image into several regions so that the contents of each region represent meaningful objects. The segmentation results can then be utilized for postprocessing stages such as object recognition. Image segmentation simplifies postprocessing stages by focusing attention on each individual segment.

Image segmentation dates back to the early 1980s. Edge detection methods such as the Canny detector [6] were widely applied for this task. Edge detection methods utilize intensity gradients to detect the boundaries of objects. However, edge detection methods usually generate edges that are not closed contours, and this causes difficulties for later processing such as object recognition. It has been shown [19] that it is hard to construct stable edge detection methods even using a multiscale theory because of the edge-linking issue across scales and multiple-thresholding issues at each scale.

Curve evolution methods [32], [33], [37] have been popular for image segmentation since the early 1990s. These methods evolve the initialized curve(s) to the boundaries of objects in an image. The evolution of the curves may be driven by image gradient information [8], [25], region information [11], [24], [35], or their combination [38]. These methods are theoretically solid and numerically stable. Moreover, these methods generate image segments enclosed by closed contours, which leads to straightforward postprocessing.

Most curve evolution methods, however, are unsupervised and are only able to segment simple images. For complicated cases, such as triple junctions and images with a cluttered background, these methods have to utilize hierarchical methods [35], and are usually computationally intense.

The utilization of prior information, therefore, seems to be necessary for curve evolution methods in complicated image segmentation applications. Several methods [13]–[17], [23], [26], [27], [34] have been proposed that utilize prior information for supervised image segmentation. These methods usually propose new variational energy functionals which integrate both the prior information and the gradient/region information in the image to be segmented. The minimizations of these functionals can lead to segmentation results. Shape priors are utilized in [13]–[15], and [27]. Both intensity priors and shape priors are applied in [26]. Natural image statistics are utilized for natural image segmentation in [23]. These methods usually work better than unsupervised methods. However, shape priors are primarily incorporated in supervised segmentation methods, and the information contained in the intensities is not always fully utilized. For example, the methods in [16] and [17] utilize the intensity histograms only, and the spatial information is ignored. Furthermore, these methods usually have initialization problems because the energy functionals have multiple local minimums. The methods are sensitive to the initial locations of the evolving curves.

The above methods tend to segment the whole image into several regions, which is challenging for images with cluttered background. On the other hand, this is not always necessary in real applications. The user may be interested in finding only the location of objects of interest. For example, only the car in Fig. 1(a) is of interest in this video tracking application. The cluttered background need not to be segmented. The task changes and can be stated as: Segment from the image the regions whose contents are similar to the car in the prior image Fig. 1(b). In this sense, image segmentation and object recognition are combined. This is the key idea of "preferential image segmentation," which means to preferentially segment objects of interests from an image and ignore the remaining portions of the image for this application.

The idea of preferential image segmentation bears some similarities to object detection methods from images [1]. These methods detect the existence and rough location of objects in an image, e.g., in [1], using a sparse, part-based object representation and a learning method. However, these methods do

Manuscript received January 19, 2008; revised October 22, 2008. First published February 27, 2009; current version published March 13, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Scott T. Acton.

The authors are with the University of Tennessee, Knoxville, TN 37996 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2008.2010202

(a) (b)

Fig. 1. Illustration of preferential image segmentation. (a) The video frame in which only the car is of interests. (b) The prior image of the car.

not provide the exact locations of the object boundaries, which is required by image segmentation. Supervised image classification methods in [3] and [36] and the image parsing method in [40] are also similar to preferential image segmentation. The image classification methods in [3] and [36] generalize the level set methods to classify an image into multiple regions by means of wavelets [3] and variational methods [36], respectively. The image parsing method [40] utilizes probabilistic methods to unify segmentation, detection and recognition. These methods, however, tend to classify/parse an whole image instead of meaningful objects. Furthermore, these methods are usually computationally intense.

A novel preferential image segmentation method is proposed in this paper using techniques from mathematical morphologies. This method is motivated by the utilization of prior information in curve evolution models. However, image topologies may provide better results for complicated cases. The proposed method utilizes a tree of shapes [29], [30] to represent the image content. This representation provides a hierarchical tree for the objects contained in the level sets of the image. The hierarchical structure is utilized to select the candidate objects from the image. The boundaries of the selected objects are then compared with those of objects selected from prior images. By means of the tree of shapes and curve matching, the proposed method is able to preferentially segment objects with closed boundaries from complicated images. It is more straightforward to utilize prior information in this way than with the curve evolution methods, and there is no initialization problem. Furthermore, the method is invariant to contrast change and translation, rotation and scale. The method has been shown to work in the presence of noise.

This paper is organized as follows. Section II provides background information on image representation using the tree of shapes and curve matching. A novel preferential image segmentation method is proposed in Section III, followed by Section IV which shows experimental results. Section V describes the evaluation of the performance of the proposed method using a large image data set. A summary and future research directions are provided in Section VI.

II. Background

Background information for the proposed method is introduced in this section. Section II-A shows how an image is represented using a tree of shapes. It also shows how a tree structure is introduced for image representation. Section II-B describes the relationship between color and geometry in natural images. Section II-C introduces the techniques of planar curve matching, which can be utilized to compare the boundaries of different objects.

A. Image Representation Using the Tree of Shapes

The tree of shapes [5], [7], [9], [20], [28], [30] represents images based on the techniques of contrast-invariant mathematical morphologies [19], [29]. This method is based on the theory of image representation using connected components of set of finite perimeters in the space of functions with weakly bounded variations (WBV), as introduced in [2], [4], and [10]. It shows that an image, if taken as a function of weakly bounded variation [4], is guaranteed to be decomposed into connected components with closed boundary [2]. This is an extension to classical methods such as [31] where an image is taken as a piecewise-smooth function.

The representation of an image using a tree of shapes utilizes the inferior or the superior of a level line to represent an object, and takes the boundary of the inferior area as the shape of the object. Therefore, only closed shapes are generated. This representation also provides a tree structure to represent the spatial relationship for the objects in an image.

For a gray image $u : \Omega \to \mathbb{R}$ with $\Omega \subset R^2$, the upper level set χ_λ of value λ and the lower level set χ^μ of value μ are defined in [7] as

$$\chi_\lambda = \left\{x \in R^2, u(x) \geq \lambda\right\} \tag{1}$$

$$\chi^\mu = \left\{x \in R^2, u(x) \leq \mu\right\}. \tag{2}$$

The above definitions have several advantages. First, they represent regions instead of curves in an image, which provide a way to handle the contents inside the regions. Second, they are invariant to the contrast changes in an image, which may be caused by the change of lighting [19], [29]. Third, closed boundaries are acquired for each upper level set or lower level set, which can be utilized for shape matching of the regions. In comparison, the level lines defined by $l_\lambda = \{x \in R^2, u(x) = \lambda\}$ usually generate open curves rather than closed curves in real images. For an image modeled as a function of bounded variation, the regions represented by level sets are connected components. Fourth, their representations are complete for images, which means that the family of the upper level sets χ_λ (or the family of the lower level sets χ^μ) is sufficient to reconstruct the image [19], [29] because of the following relationship [7]:

$$u(x) = \sup\{\lambda | x \in \chi_\lambda\} = \inf\{\mu | x \in \chi^\mu\}. \tag{3}$$

Note that the geometrical inclusion holds for the level sets. The family of upper (lower) level sets is decreasing (increasing) because [7]

$$\lambda \leq \mu \Rightarrow \chi_\lambda \supset \chi_\mu \quad \text{and} \quad \chi^\lambda \subset \chi^\mu. \tag{4}$$

The nesting of level sets provides an inclusion tree for an image. The inclusion tree from the family of upper level sets and the tree from the family of lower level sets, however, can be different if the connected components are directly utilized. The

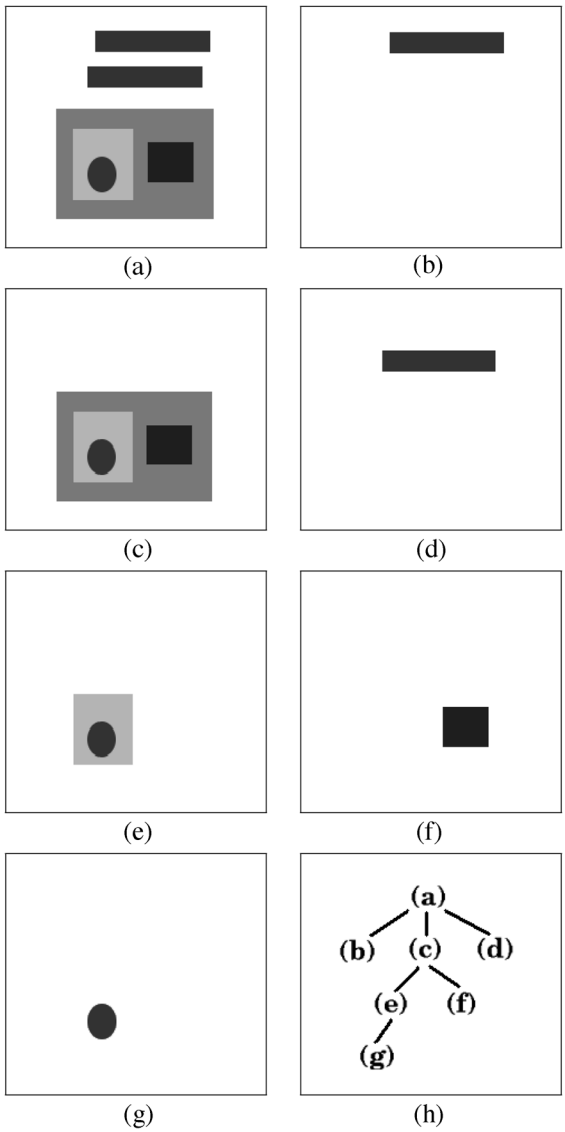


Fig. 2. Illustration of the tree of shapes. (a) The original image, which locates at the root of the inclusion tree. (b), (c), (d) Shapes in the first layer of the tree of shapes. (e), (f) Shapes in the second layer of the tree of shapes. (g) Shapes in the third layer of the tree of shapes. (h) The structure of the tree of shapes.

concept of "shape" is introduced to generate a unique inclusion tree for an image. A shape is defined as the connected components of a level set and the holes inside them. Fig. 2 shows an example of the tree of shapes generated for a piecewise-constant image. A tree of shapes shown in Fig. 2(h) is constructed for the image in Fig. 2(a). The whole image acts as the root of the tree, which locates at the top level. The shapes in the same level are spatially disjoint in the image. The shapes in the lower level are spatially included in the shapes in the next higher level. The tree of shapes, therefore, provides a natural way to represent the spatial relationships between the shapes in the image.

It is straightforward to find upper level sets and lower level sets in an image by thresholding. A tree of shapes can be further constructed by the nesting of level sets. However, this method is computationally intense. The fast level lines transform (FLLT) [29], [30] provides a faster way to construct a tree of shapes. FLLT is a pyramidal algorithm based on region growing. It reduces the computational burden from $O(N^2)$ to $O(N \log N)$ where N represents the number of pixels in an image. FLLT is implemented in the MEGAWAVE package (http://www.cmla.ens-cachan.fr).

B. Color and Geometry in Mathematical Morphologies

The total order (or lexicographical order) proposed in [12], [21] is utilized in the proposed method to compare the color vectors in a color image. Let $\mathbf{C_1} = (Y_1, S_1, H_1)$ and $\mathbf{C_2} = (Y_2, S_2, H_2)$ represent the color vectors in two pixels in a color image. The meanings of Y_1, S_1, and H_1 differ in different color models. The total order provides a way to compare these color vectors. It is defined as follows:

$$\mathbf{C_1} < \mathbf{C_2} \quad \text{if} \quad \begin{cases} Y_1 < Y_2 \text{ or} \\ Y_1 = Y_2 \text{ and } S_1 < S_2 \text{ or} \\ Y_1 = Y_2 \text{ and } S_1 = S_2 \text{ and } H_1 < H_2. \end{cases} \tag{5}$$

The definition in (5) bears some similarity with the concept of conditional expectations shown in Section II-B when Y_1 represents the gray level in the image.

Y_1 actually represents the gray level in the color image in the proposed method. An improved $HLS(IHLS)$ space [22] is utilized for the color model. The color in every pixel is represented with three channels (Y, S, H), which corresponds to the gray level, saturation and hue respectively. The $IHLS$ space, compared to other spaces such as HLS and HSV, has the property of a "well-behaved" saturation coordinate. The $IHLS$ space always has a small numerical value for near-achromatic colors, and is completely independent of the brightness function. For a pixel with color (R, G, B) in the RGB space, and the corresponding pixel with color (Y, S, H), the transformation from the RGB space to the IHLS space is

$$Y = 0.2126R + 0.7152G + 0.0722B \tag{6}$$

$$S = \max(R, G, B) - \min(R, G, B) \tag{7}$$

$$H = \begin{cases} 360^\circ - H', & \text{if } B > G \\ H', & \text{otherwise} \end{cases} \tag{8}$$

where

$$H' = \frac{R - 0.5G - 0.5B}{(R^2 + G^2 + B^2 - RG - RB - BG)^{\frac{1}{2}}}. \tag{9}$$

The proposed method uses the transformation and the total order in (5) to extract the shapes and build the tree of shapes for color images. The inverse transformation from the IHLS space to the RGB space is not utilized here. Refer to [22] for more details.

Fig. 3 illustrates the construction of a tree of shapes for a color image. Fig. 3(a) displays a color image of interest. The color image is synthesized so that its gray version contains no shape information. Using the IHLS color space (6)–(9) and the total order (5), a tree of shapes is built for the color image, whose boundaries are shown in Fig. 3(b).

C. Planar Curve Matching

The method in [28] defines the shape of a curve as a conjunction of shape elements and further defines the shape elements as any local, contrast invariant and affine invariant part of the

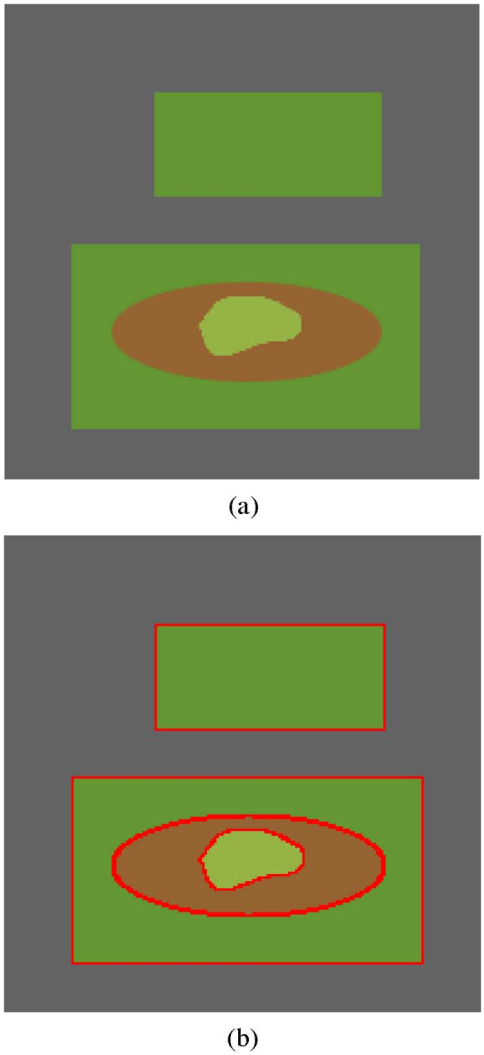


Fig. 3. Illustration of tree of shapes for a color image. (a) A color image of interests. (b) Boundaries of shapes built for (a) using the total order defined in (5).

curve. These definitions are oriented to provide invariance to noise, affine distortion, contrast changes, occlusion, and background.

The shape matching between two images are designed as the following steps.

1) Extraction of the level lines for each image. The level set representations (1), (2) are utilized here for the extraction. The level lines is defined as the boundaries of the connected components as shown before.
2) Affine filtering [28] of the extracted level lines at several scales. This step is applied to smooth the curves using affine curvature deformation to reduce the effects of noise.
3) Local encoding of pieces of level lines after affine normalization [28]. Both local encoding and affine normalization are designed for local shape recognition methods. This step will help to deal with occlusions in real applications.
4) Comparison of the vectors of features of the images. Euclidean distance is utilized to compare the feature vectors.

The performance of curve matching between two curves is calculated after affine filtering, curve normalization and local encoding. Suppose C_1 and C_2 are two curves for matching, and S_1 and S_2 are pieces from C_1 and C_2, respectively. The performance to take S_1 and S_2 as the matching between C_1 and C_2 is

$$\text{Score} = \frac{l_1 \times l_2}{L_1 \times L_2} \tag{10}$$

where $l_1 = \text{arclength}(\text{S}_1)$, $l_2 = \text{arclength}(\text{S}_2)$, $L_1 = \text{arclength}(\text{C}_1)$ and $L_2 = \text{arclength}(\text{C}_2)$. The maximum score over all possible matching pieces is taken as the matching between the curves C_1 and C_2. Refer to [28] for details of the mathematical definitions and the implementation issues. This method is implemented in the MEGAWAVE package (http://www.cmla.ens-cachan.fr), which is open-source and can be utilized directly.

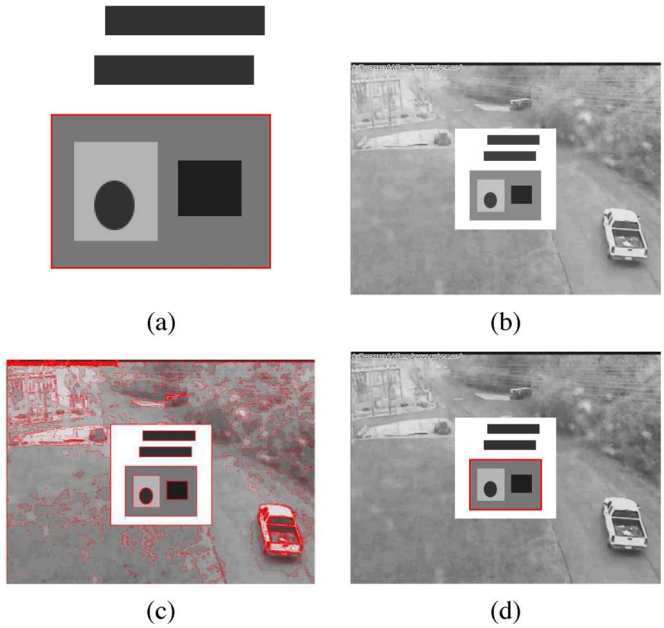


Fig. 4. Illustration of of preferential image segmentation using the tree of shapes. (a) The object of interest in the prior image. (b) The image to be segmented. (c) A simplified tree of shapes of the whole image in (b). (d) Segmentation results using the proposed preferential image segmentation. Image size: 460 $*$ 612. CPU time: 1.70 s.

III. Preferential Image Segmentation

A preferential image segmentation algorithm is proposed in this section. The proposed algorithm is illustrated using an example for clarification. Suppose we need to preferentially segment objects similar to the one indicated in Fig. 4(a) from the cluttered image shown in Fig. 4(b). As illustrated in Section II-A, the tree of shapes provides a natural way to represent the spatial relationship of the shapes in an image. The tree of shapes, therefore, is a good candidate tool to utilize the intensity information in the prior image for preferential segmentation. A naturally intuitive idea is to construct the trees of shapes for both images and then find the node in the tree of image in Fig. 4(b) whose properties are similar to the node in Fig. 2(c).

The tree of shapes, however, usually generate large number of shapes, especially for complicated images. For example, 3832 shapes, as shown in Fig. 4(c), are contained in the tree of shapes for the whole image in Fig. 4(b). Note that every closed red

curve corresponds to the boundaries of a shape in the tree of shapes in Fig. 4(c).

The boundaries of shapes provide valuable information for the similarities between shapes. However, comparison of all the boundaries corresponding to shapes in the image to be segmented would be computationally intense. It is necessary to narrow down the candidate shapes from the image to be segmented for each shape from the prior image before their boundaries can be compared. The intensity information contained in the shapes provides good measures for the similarities between shapes in real images. The intensity information by means of the tree of shapes [e.g. Fig. 2(h)] includes the following features.

1) The number of objects Nd contained directly in the shape, which corresponds to the number of direct children of the shape in the tree. $Nd = 2$ for the indicated shape in Fig. 4(a).
2) The total number of objects Nt contained in the shape, which corresponds to the total number of children below the shape in the tree. $Nt = 3$ for the indicated shape in Fig. 4(a).
3) The relative area change A between the shape and its direct children. Suppose the area of the shape is S, and the areas of its direct children are S_i, where $1 \leq i \leq Nd$, the relative area change is then defined as

$$A = \prod_{i=1}^{N_d} \frac{S_i}{S}. \tag{11}$$

4) The rough similarities R of the boundaries of the shapes, which can be represented as the ratio of the circumferences C squared to the area S of the boundaries of the shapes [18], i.e., $R = C^2/S$.

These features for two shapes should be very close if they match. Exact matching may not be achieved because of the differences between shapes, the effects of noise, lighting changes and cluttered background. Thresholds should be set for coarse matching to affect a rapid exclusion of most candidate shapes. For example, set the threshold $\mathrm{Nd_{thres}} = 1$, $\mathrm{Ratio_{thres}} = 0.3$. A shape is considered to roughly match the prior match if $\mathrm{abs(Nd_{prior} - Nd)} \leq \mathrm{Nd_{thres}}$, $\mathrm{abs(Nt/Nt_{prior} - 1)} < \mathrm{Ratio_{thres}}$, $\mathrm{abs(A/A_{prior} - 1)} < \mathrm{Ratio_{thres}}$, $\mathrm{abs(R/R_{prior} - 1)} < \mathrm{Ratio_{thres}}$. These thresholds may be adjusted for different applications.

The number of candidate shapes decreases substantially by means of the intensity features extracted from the tree of shapes. In the case of Fig. 4(b), the feature Nd decreases the number of candidate shapes from 3832 to 977; the feature Nt decreases the number from 977 to 83; the feature A decreases the number from 83 to 1; the feature R retains this candidate. The candidate shape left over matches the prior shape, as shown in Fig. 4(d). The process takes 1.70 s.

In most cases, however, the candidate shapes will decrease from thousands to tens, but not 1. Curve matching as introduced in Section II-C is then performed on the boundaries of the resulting candidate shapes and the prior shapes. The candidate shape which best matches the prior's curve is taken as the preferential segmentation result.

In summary, the proposed method models an image with a function of bounded variation and utilizes a tree of shapes for image representation. This method encodes the prior information for preferential segmentation as a tree of shapes. The method is invariant to translation, rotation and scale transformations because both feature extraction and boundary matching are invariant to these transformations.

The proposed method bears similarities with template matching. Both methods utilize prior information to locate objects from images. However, these methods are different in various aspects. First, the proposed method models an image as a function of bounded variation, while template matching usually models an image as a continuous function. Second, the proposed method utilizes level sets for image representation and represents spatial relationship by a tree structure, while template matching represents an image as pixels by means of sampling theorem. Third, the proposed method handles invariance to rotation, scale and translation automatically, while template matching needs to handle these invariances explicitly. Fourth, the proposed method is easy to extend to segmentation by parts, while template matching is hard for occlusions. More importantly, the proposed method segments a most similar object from an image, which combines segmentation and recognition. A template matching method, on the other hand, only recognizes the location, orientation and scale information of a known object. There is no segmentation in template matching. Experimental results for this comparison are shown in the next section.

IV. Experimental Results

Experimental results from the proposed method are provided in this section. The proposed method is implemented using the MEGAWAVE package (http://www.cmla.ens-cachan.fr). The program runs on a computer which has two Intel(R) Pentium(R) 3.2 GHz CPUs, 1-Gb RAM, and runs the Red Hat Enterprise Linux operating system. The CPU times given in this paper are the sums of system CPU times and user CPU times. The system CPU time is usually very small, typically 0.01–0.08 s.

Two implementation issues need to be clarified here. First, the representative shapes are manually selected from the inclusion tree of the prior image. The "llview" function of the MEGAWAVE package is utilized to visualize the inclusion tree for shape selection. This is acceptable because the operation only need to be performed once. Automatic selection is the focus of current research. Second, interpolation methods are utilized in the construction of the inclusion tree for high precision. This may generate many shapes whose areas are very close (e.g., 95%–98%). These shapes may correspond to only one object in the image. A sensitivity measure is utilized to handle this issue. The shape will not be considered as an object until the ratio of its area to that of its precedents falls below a certain threshold (e.g., 80%).

Fig. 5 shows the preferential image segmentation of a car from an image with cluttered background. For the prior image Fig. 5(a), the representative shape is selected as shown in Fig. 5(b). A very similar shape is selected from the image Fig. 5(c),

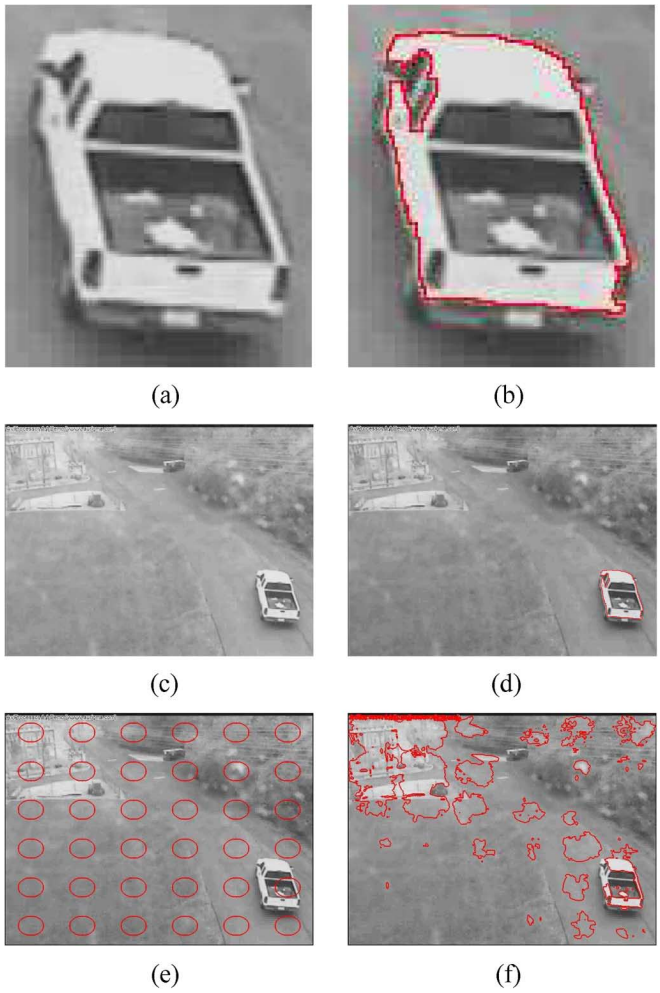


Fig. 5. Preferential image segmentation of a car using the tree of shapes. (a) The prior image (image size: 460 $*$ 612). (b) The object of interest in the prior image. (c) The image to be segmented. (d) Segmentation results using the proposed preferential image segmentation. The region enclosed by the red line corresponds to the segmentation result similar to (b). CPU time: 1.53 s. (e) The initialized curves for the Chan–Vese curve evolution. (f) The intermediate segmentation result using curve evolution after 800 iterations. CPU time: 712 s.

as shown in Fig. 5(d). Curve matching is not necessary here because only one candidate is left over. The process takes 1.53 s. For comparison, the results in Fig. 5(d) are also achieved by means of template matching, whose implementation by the authors takes 192.57 s, which is very time consuming compared to the proposed method (126×). Fig. 5(f) shows the intermediate segmentation results using the Chan–Vese curve evolution method [11] without using any prior information for the initialization shown in Fig. 5(e). It can be seen that the curve evolution method segments the brighter area of the image, which includes parts of the car, but it cannot segment the whole car. The curve evolution method is computationally intense, requiring almost 12 min in this case. Furthermore, the post processing for the segmentation results is challenging because of the cluttered background. The curve evolution methods based on prior information may provide better segmentation results, but it can be expected that they would be also computationally intense. Thorough comparisons are a subject of future research.

Fig. 6 shows the performance of the proposed method for a MRI brain image and a car logo. The prior shape enclosed in the red line in Fig. 6(a) is used to segment the image in Fig. 6(b), which also contains the segmentation results. Only 0.61 s was required for the segmentation of an image of size 218 $*$ 282. Similarly, the prior shape in Fig. 6(c) is utilized for the segmentation of Fig. 6(d), which also contains the segmentation results. This took 4.98 s for the segmentation of an image of size 600 $*$ 450.

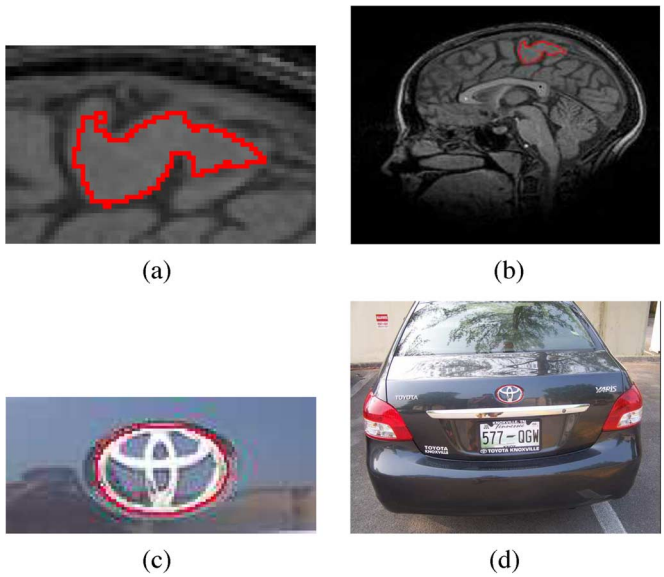


Fig. 6. More examples for preferential image segmentation. (a) The prior image for a medical image. (b) Segmentation results of using the prior shape from (a). Image size: 218 $*$ 282. CPU time: 0.61 s. (c) The prior image for a car logo. (d) Segmentation results using the prior shape from (c). Image size: 600 $*$ 450. CPU time: 4.98 s. The regions enclosed by the red lines corresponds to the segmentation results.

The rotation invariance of the proposed method is illustrated in Fig. 7. The shape shown in Fig. 7(a) is utilized as the prior for preferential image segmentation in Fig. 7(b). The candidate shapes resulting from intensity features are shown in Fig. 7(c) Both shapes are similar to the one in Fig. 7(a). Fig. 7(d) shows the final result, which agrees with the ground truth.

The proposed method is also invariant to scale change, as shown in Fig. 8. The object shown in Fig. 8(a) (size: 147 $*$ 113) is extracted from Fig. 8(d), and is then scaled up to Fig. 8(b) (size: 195 $*$ 150), which is utilized as the prior for preferential segmentation in Fig. 8(d). Fig. 8(c) shows the prior shape for Fig. 8(b). Fig. 8(d) shows the segmentation results. The object is found using a scaled prior image, which demonstrates scale invariance.

Contrast invariance and intensity invariance of the proposed method are demonstrated in Fig. 9. Fig. 9(a) shows the object extracted from the image to be segmented. The image contrast is increased, as shown in Fig. 9(b). The object is successfully segmented from Fig. 9(d) using the selected prior in Fig. 9(c). The intensity levels are increased for the image in Fig. 9(e). The results in Fig. 9(g) show that the object has been successfully segmented by means of the prior shape in Fig. 9(f).

The results shown in Fig. 10 show that the proposed method works in the presence of noise. The prior shape in Fig. 10(a) are utilized for preferential segmentation. Fig. 10(b) shows the segmentation results for a noisy image. Fig. 10(c) displays the

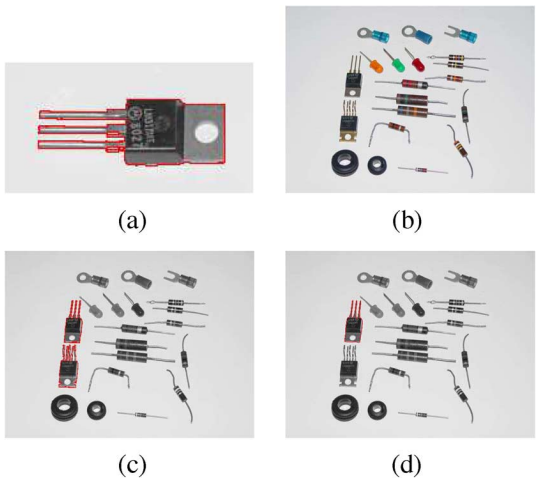


Fig. 7. Illustration of rotation invariance of the proposed method. (a) The prior image with object of interests. (b) The image to be segmented. Image size: 426 $*$ 568. (c) Candidate shapes from the intensity prior. (d) Segmentation result. The region enclosed by the red line corresponds to the segmentation result. CPU time: 2.74 s.

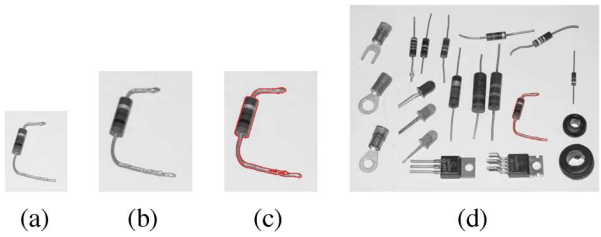


Fig. 8. Illustration of scale invariance of the proposed method. (a) The object extracted from the image to be segmented. Size: 147 $*$ 113. (b) The rescaled image from (a). Size: 195 $*$ 150. (c) Prior shape selected from (b). (d) Segmentation result. The region enclosed by the red line corresponds to the segmentation result. CPU time: 2.73 s.

results for the same noisy image after diffusion. The proposed method works in both cases, though the shapes of the segmented regions are slightly different from the prior object's shape due to the effects of noise. It can also be expected that the method may fail in the case of heavy noise because noise affects the intensity features and curve matching.

A collection of 26 images is utilized to test the performances of the proposed method. The objects in 16 images are successfully segmented based on the prior image in Fig. 8(c). Fig. 11 shows the segmentation results of four exemplar images. The results show that the proposed method works for complicated images with shadows.

The proposed method is applied to face images in Fig. 12. Twenty face images (image_0001.jpg–image_0020.jpg from the Caltech face image database, which is available at http://www.vision.caltech.edu/archive.html) are tested. The face in image_0018.jpg is extracted for the prior image. Three images with different backgrounds are segmented successfully, as shown in Fig. 12(b)–(d). Expression changes in the face, such as winking eyes, and opening and closing of eyes or mouth, may cause sufficient changes in the face images to introduce difficulties for preferential image segmentation using the proposed method. Segmentation by parts may be a good solution to this issue.

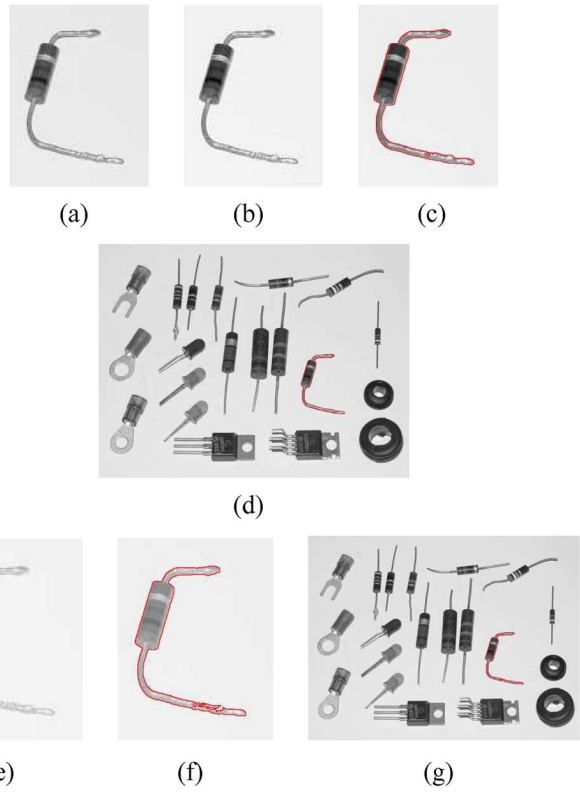


Fig. 9. Illustration of contrast and lighting invariance of the proposed method. (a) The object extracted from the image to be segmented. (b) The image with higher contrast. (c) Prior shape selected from (b). (d) Segmentation result using (c). Image size: 426 $*$ 568. CPU time: 1.61 s. (e) The image with higher brightness. (f) Prior shape selected from (e). (g) Segmentation result using (f). Image size: 426 $*$ 568. CPU time: 2.57 s.

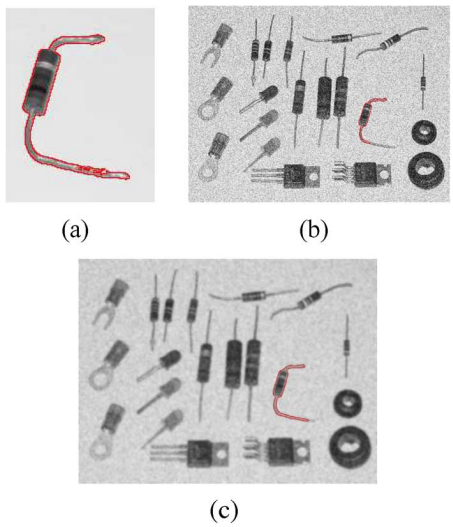


Fig. 10. The proposed method works to the effects of noise. (a) The prior shape. (b) Segmentation results for a noisy image (Size: 568 $*$ 426). CPU time: 4.91 s. (c) Segmentation results for the noisy image in (b) after diffusion. CPU time: 3.68 s.

V. Performance Evaluation of Preferential Image Segmentation Using the Tree of Shapes

The previous sections show the theoretical background, algorithm details and experimental results of the proposed preferential image segmentation method. The results demonstrate that

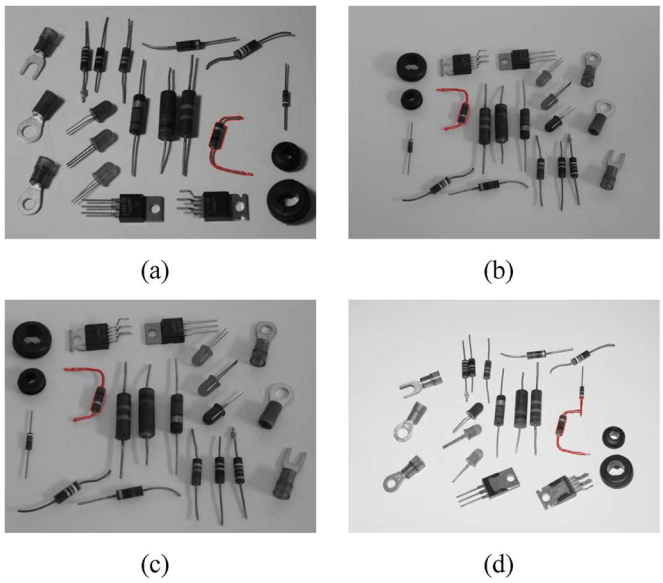


Fig. 11. Preferential image segmentation for a collection of images. (a) Segmentation results. CPU time: 3.62 s. (b) Segmentation results. CPU time: 3.09 s. (c) Segmentation results. CPU time: 4.26 s. (d) Segmentation results. CPU time: 2.03 s. Image size: 460 $*$ 612.

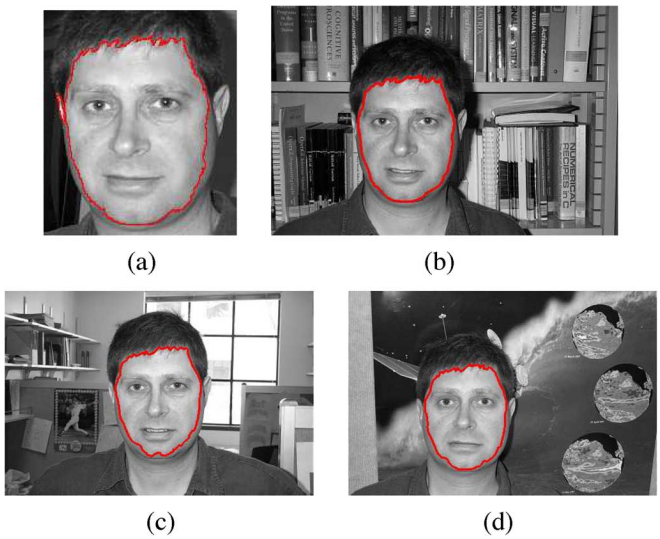


Fig. 12. Preferential image segmentation of face images (from http://www.vision.caltech.edu/archive.html) using the Tree of Shapes. (a) The prior image with selected shape. (b) Segmentation results (image size: 296 $*$ 448). CPU time: 2.43 s. (c) Segmentation results for another face image (size: 296 $*$ 448). CPU time: 2.26 s. (d) Segmentation results (size: 296 $*$ 448). CPU time: 2.59 s.

the proposed method is efficient, invariant to contrast changes and similarity transformations, and applicable to various applications. A performance evaluation of the proposed method using a large set of natural images is presented in this section. The results here will give a better indication of the performance of the proposed methods in real applications.

This section presents the procedure used for performance evaluation and the corresponding results. The experimental results are also provided in this section. An introduction to the image dataset is described in Section V-A. The way to select prior shapes from the training images is illustrated in Section V-B. This is followed by the details of the experimental results in Section V-C.

A. Introduction to the Image Dataset

The image dataset contains a training set of 26 images and a test set of 95 images. These images are taken under different conditions of lighting, backgrounds, shadows, displacements, and rotations, and examples of close contact between objects and strong occlusions. Every training image contains 20 objects whose prior information is known. The number of objects varies in the test images. The total number of objects in the 95 test images is 520. Fig. 13(a) shows the indices of the prior objects in a training image. Fig. 13(b)–(d) shows examples of the training images with good lighting and clear background. Fig. 13(e) and (f) are examples of the training images with light shadows. Fig. 13(g) and (h) are examples of the training images with strong shadows. Occlusions and rotations can be seen in Fig. 13(c) and (d).

The test images are complicated by occlusion, textured background, strong shadow and close contact. Fig. 14 shows examples of the images to be segmented in the test set. Note that the training images have no textured backgrounds, which introduces more challenges to the segmentation task. The test images are categorized into several classes: clear background, cluttered (or textured) background, light shadow, strong shadow, close contact, occlusion, separated. The first two classes represent the background information; the middle two classes show the lighting conditions; the last three attributes represent the spatial relationships between objects. These categories may overlap. Among the 95 test images, 47 test images have clear backgrounds, e.g., Fig. 14(a), (d), (g), (h); 33 test images have cluttered background, e.g., Fig. 14(b), (c), (e), (f); seven test images have light shadows, e.g., Fig. 14(f); 34 test images have strong shadows, e.g., Fig. 14(g), (h); 30 test images have close contacts among their objects, e.g., Fig. 14(e); 41 images have occlusions, e.g., Fig. 14(b), (c), (d), (f), (g); 24 test images have spatially separated objects. These categories will be analyzed separately to determine the influences of these factors on the segmentation results.

B. Prior Selection

The experiments are designed to utilize the prior information provided in the training set to segment the objects in the test set. The priors in the images of the training set are first manually segmented. An interface for the prior selection has been developed with the help of the CImg library [39]. The program will preferentially select the shape which contains the point left-clicked by the mouse and which has at least one direct child. The selected shape will be removed by right clicking the mouse if the selected object is not satisfactory. Fig. 15 shows the procedure used to select object 18 from the training image Quarter753.bmp. The user clicks the pixel under the red cursor, located in the left part of the training image, and the boundary of the selected object is shown in the right part.

Fig. 16 shows several examples of prior objects selected from the training images. Fig. 16(a) and (b) shows prior objects selected from images with no shadows or occlusions. Fig. 16(c)–(f) shows prior objects selected in the presence of light or strong shadows, which degrades the prior selection process. Parts of the prior objects are selected when the whole

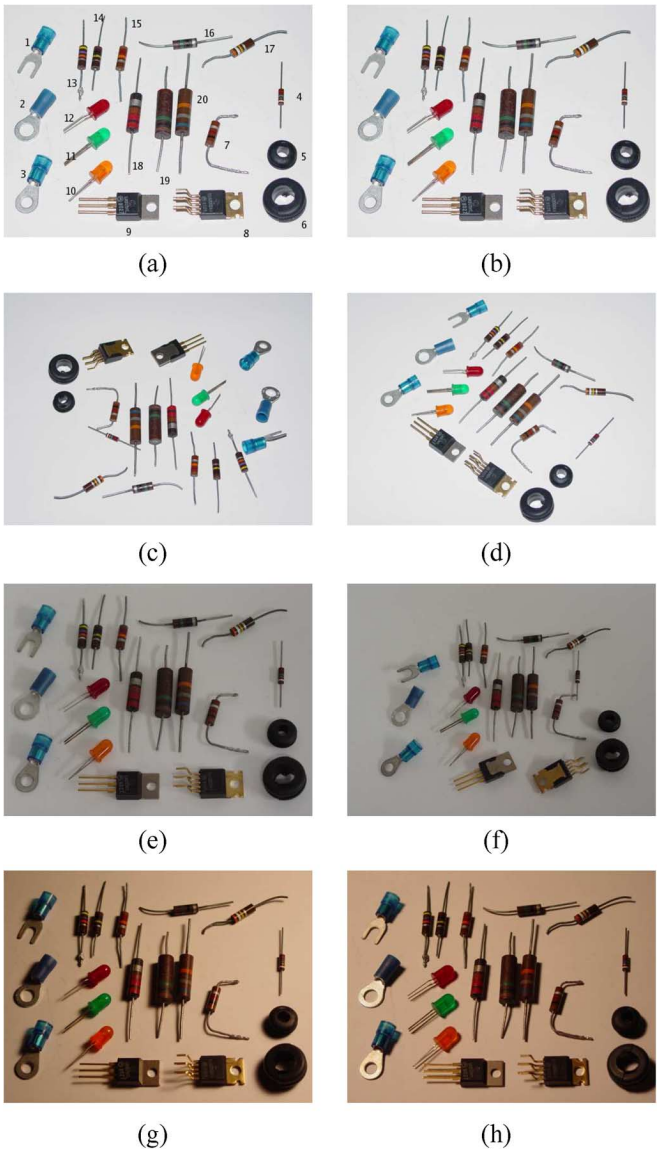


Fig. 13. Examples of prior images in the training set. (a) The indices of objects in the prior image. (b)–(h) Examples of prior images with different background, lighting, shadows, poses and occlusions. The tree structures for thess images contain 26474, 16755, 17035, 60027, 43190, 65454, 56250 shapes, respectively, in the implementation.

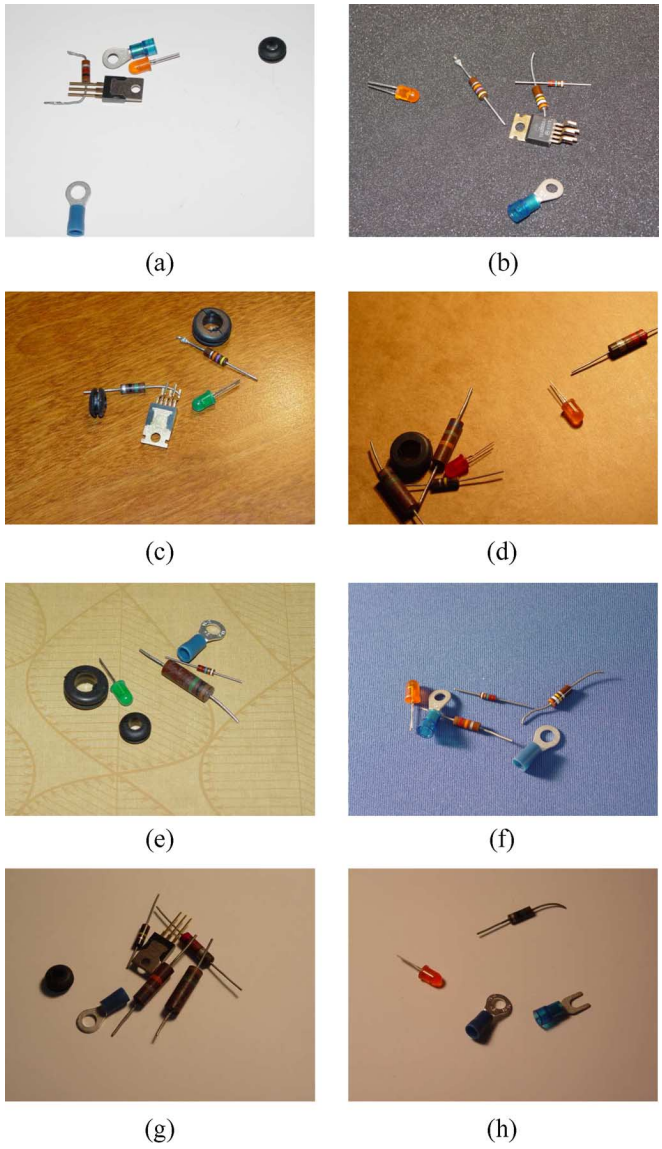


Fig. 14. Examples of images to be segmented in the test set. (a)–(h) Examples of test images with different background, lighting, shadows, poses, and occlusions.

objects are difficult to select because of the presence of shadow or occlusion, as shown in Fig. 16(g) and (h).

The manual selection procedure is offline and identifies the priors from the training images. The prior information utilized in the proposed method contains only the number of direct children N_d, the total number of children N_t, the relative area change A, the rough similarities R, and the coordinates of the pixels on the boundaries of the object. This information is stored in a sequential text file on the hard drive. The text file will be read into memory when the test images are processed.

As an example, the extracted information for the shape prior selected in Fig. 15 (object 18 in Quarter753.bmp) is displayed as follows:

```
1 1 52 1 0 116.199 525 362717
2 0 252 205 ...
```

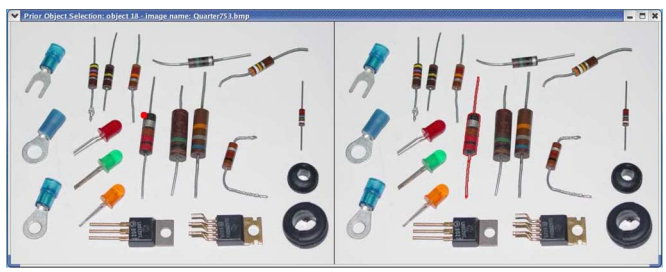


Fig. 15. Illustration of shape prior selection in training images.

Each of the numbers is stored as a line in the record. These numbers are displayed as a line for the convenience of visualization. The first number (1) represents the type of the level set. Its value is set to 1 if the selected shape prior is a upper level set; otherwise 0. The second number (1) represents the direct children of the selected shape prior. The third number (52)

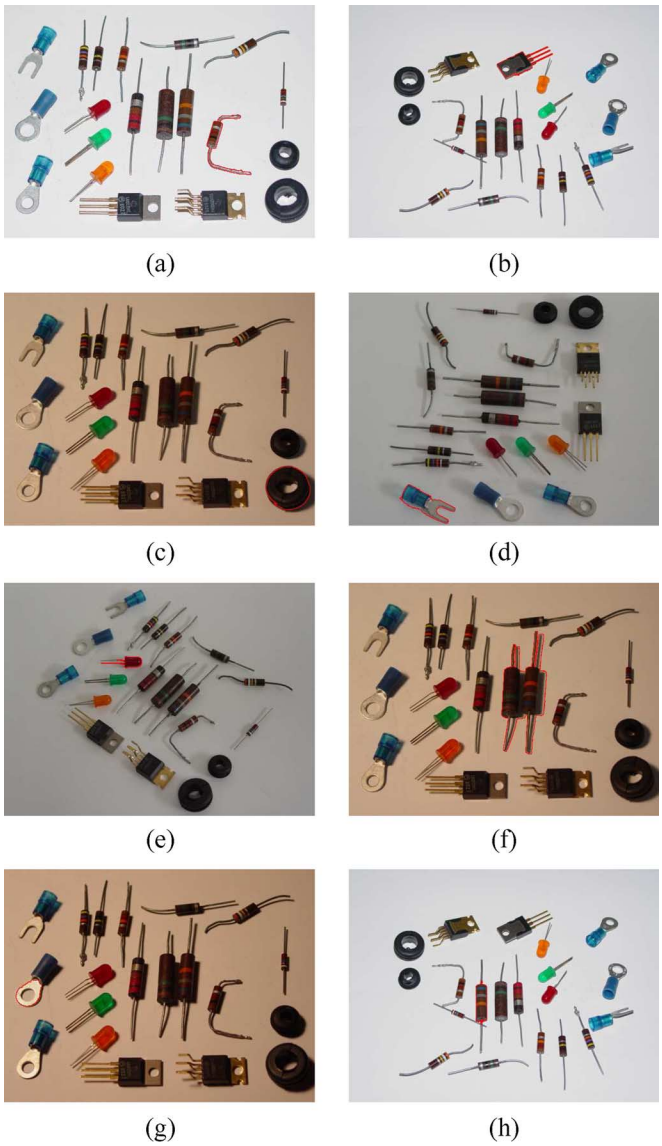


Fig. 16. Examples of prior objects selected from the training images. (a)–(h) Examples of prior objects selected from the training images images with different background, lighting, shadows, poses, and occlusions.

stores the number of total children. The forth number stands for the ratio of the relative area change, defined in (11). The fifth number (0) represents the depth of the current shape. The sixth number (116.199) represents the boundary-squared-to-area ratio of the selected shape. The numbers left over are utilized for boundary matching. The seventh number (525) shows the size of the boundary for the selected shape. The eighth number (362717) represents the maximum size available for the boundary storage. The ninth number (2) is the dimension of the boundary pixels, which is usually 2. The tenth number in the file represents the data for program use. The eleventh number and the twelfth number represents the coordinates of a pixel in the boundary. The rest of the file shows the coordinates of all other pixels contained in the boundary of the selected shape.

C. Experimental Results

In the prior selection stage, one object is selected from the training images at a time. Thus, $26 \times 20 = 520$ prior objects

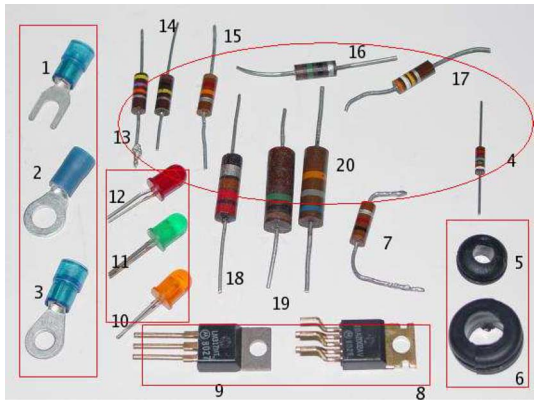


Fig. 17. Loose criteria for false alarm evaluation. (a) Groups of objects which are taken as inseparable by the visual evaluation criteria.

are obtained. The prior information contained in every object is utilized to segment each of the 95 test images. Thus, $26 \times 20 \times 95 = 49,400$ experiments are performed to evaluate the performance of the proposed methods. In the detection of object 1 from the test image Quarter779.bmp, for example, all 26 priors for object 1 are utilized one at a time, and the one that matches best is recorded as the segmentation result.

Human Visual Evaluation: The most reliable way to evaluate the experimental results is by human eyes. 456 out of 520 objects (87.7%) in the 95 test images have been correctly segmented according to this evaluation procedure. Less than 20 objects were falsely segmented for those 1380 objects that do not exist in test images using a human evaluation criteria. This is a false alarm rate of about 1.4% using that evaluation.

However, results from human evaluations are subjective. Users may get different results from their evaluations because different criteria are utilized. A loose criterion is utilized here for an illustration. Fig. 17 shows the groups of objects which are taken as inseparable by the visual evaluation criteria. For example, if object 1 is not shown in an image, but the program segments object 2 (or object 3) as the result, it is taken as false alarm. The accuracy may increase to above 90%, but at the same time the false alarm rate rises 24% (331 out of 1380). This shows that the false alarm rate of the proposed algorithm lies between 1.4% and 24%, depending on the criteria. Similarly, the accuracy lies between 87.7% and 100%.

Figs. 18 and 19 show several segmentation results. The left column in Fig. 18 shows the prior object, while the right column shows the segmentation result. Fig. 18(b) shows an example where the proposed method works very well for good lighting with no occlusions. Fig. 18(d) shows an example where an object can be segmented by parts in cluttered background. Fig. 18(f) shows that an object can be segmented when it is placed in different poses. Fig. 18(h) shows that the proposed method may also work well in the presence of strong shadows.

Fig. 19 shows several segmentation results from challenging problems. The left column in Fig. 19 shows the prior object, while the right column shows the segmentation results. The results in Fig. 19(b) show that close contacts between objects may introduce difficulties for preferential segmentation. Fig. 19(d) shows the performance of the method in textured background

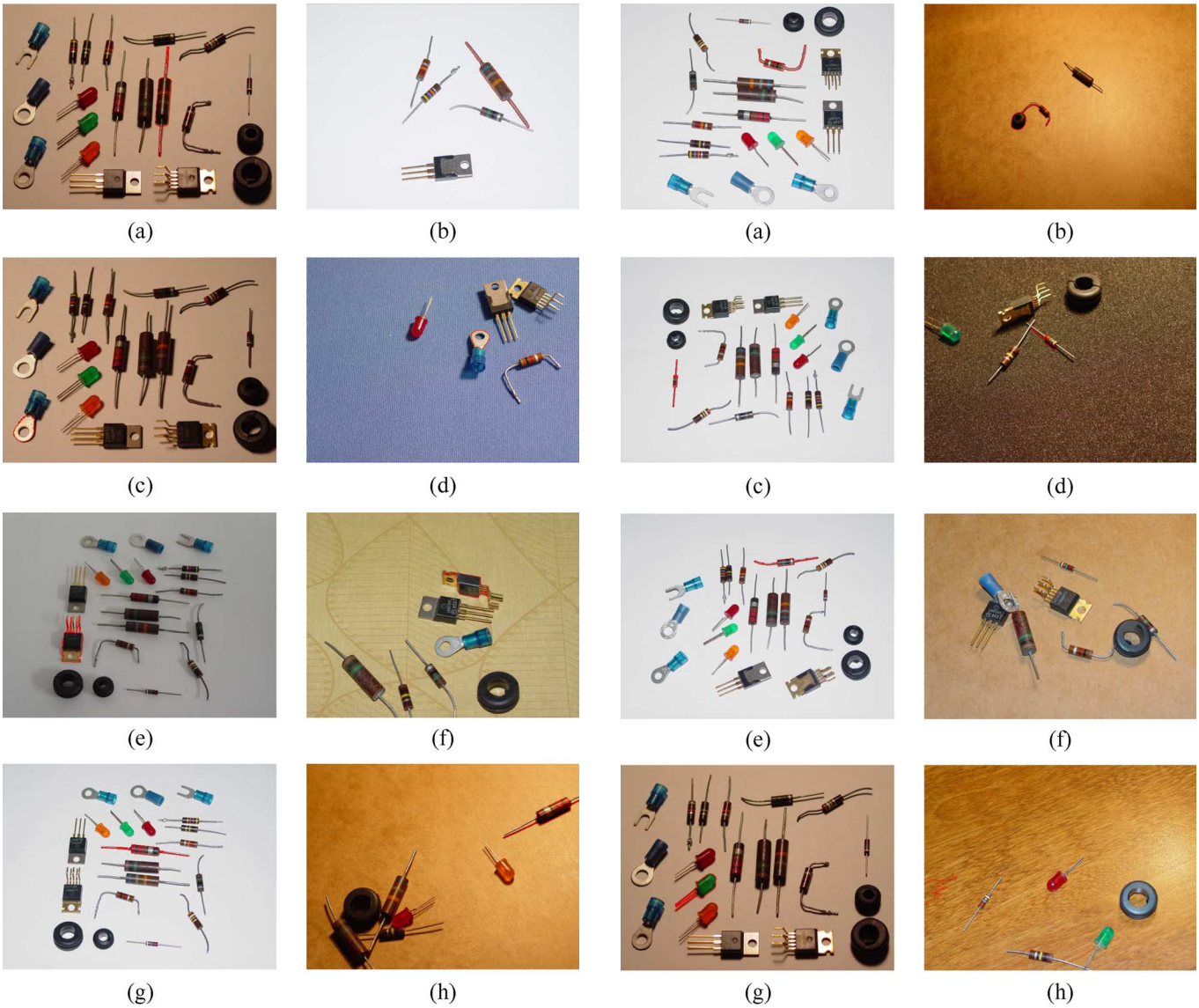


Fig. 18. Examples of segmentation results. (a), (c), (e), (g). Prior objects in the training images. (b), (d), (f), (h). Segmentation results of the test images using the priors in (a), (c), (e), (g), respectively.

Fig. 19. Examples of segmentation results. (a), (c), (e), (g). Prior objects in the training images. (b), (d), (f), (h). Segmentation results of the test images using the priors in (a), (c), (e), (g), respectively.

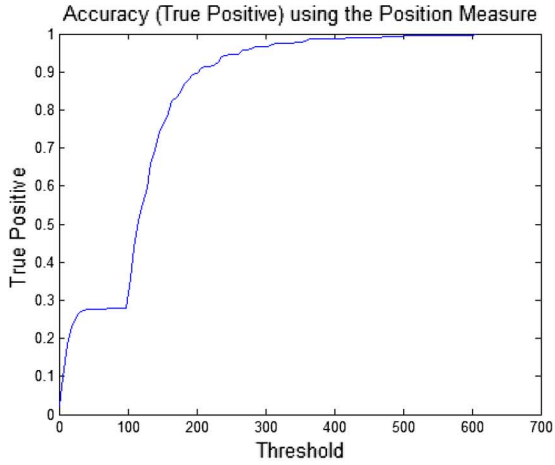
and strong shadows. Because of occlusion, object 16 cannot be segmented in Fig. 19(f). Fig. 19(h) shows that a cluttered background can cause preferential segmentation to fail.

The previous discussion shows the necessity for automatic and objective evaluation. Human visual performance evaluation is subjective rather than objective. It also requires human intervention and is, therefore, not automatic, which is a big disadvantage for large image datasets or autonomous operations. Automatic performance evaluation needs a quantitative evaluation of the match quality between the prior object [e.g., Fig. 19(a)] and the corresponding segmented object [e.g., Fig. 19(b)]. An automatic performance evaluation using the location center is presented below.

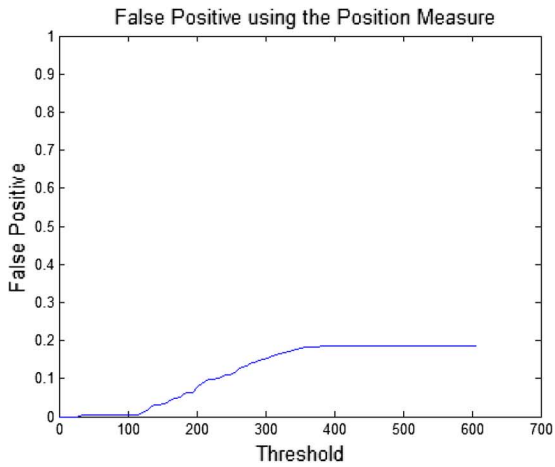
Automatic Performance Evaluation Using the Location Center: The locations of segmented objects provides a way to evaluate the performance of the proposed method. If the center of a segmented object is close to the center of visual observation, then preferential segmentation may be considered as successful. Otherwise it may be considered a failure. This method does not give a thorough evaluation because different segments may have the same center, and that the content and boundary information are not utilized. However, it provides a way to evaluate the locations of the segmented objects.

In the implementation, the locations of the objects are first visually determined first. The locations of the segments are calculated by the program, and the Euclidean distances between the corresponding locations are calculated. The minimum of the distances between all priors is taken as the final result. If an object does not exist in an image, or the program does not segment anything for an object, the location of that object is taken as (0, 0). Therefore, this evaluation method produces a small false alarm (false positive) rate and a high accuracy (true positive) rate relative to other evaluation strategies.

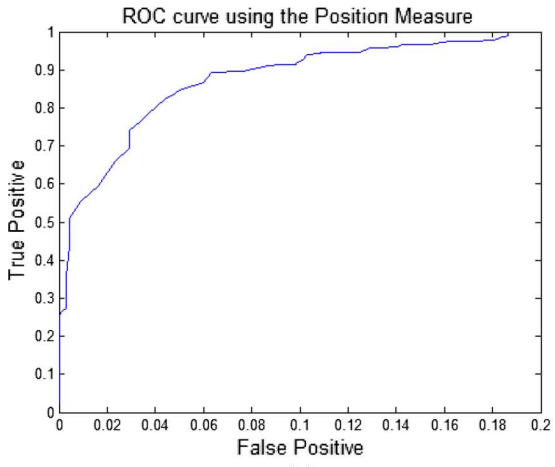
Fig. 20 shows the figures of the true positive rate [Fig. 20(a)] and the false positive rate [Fig. 20(b)], with respect to the distance threshold in pixels, as well as the receiver operating char-



(a)



(b)



(c)

Fig. 20. Performance evaluation with respect to the distance threshold. (a) Accuracy rate (true positive rate). (b) False alarm rate (false positive). (c) ROC curve for the evaluation.

acteristics (ROC) curve (with the true positive rate as the vertical axis and the false positive rate as the horizontal axis) [Fig. 20(c)]. One issue about the ROC curve is that the maximum false positive rate is less than 20%. A trade-off between detection capability and error rates can be achieved by choice of detection threshold. If a 91% detection rate is desired, a threshold of 206.47 is chosen, yielding a false positive rate of 8.7% and a failure-to-detect rate of 9.0%.

VI. Summary and Future Research

A novel preferential image segmentation method is proposed in this paper. The method utilizes both the intensity and shape prior information by means of the tree of shapes and boundary matching. It is invariant to contrast change and similarity transformations such as scale, rotation and translation. Experimental results show that the proposed method is able to preferentially segment objects from complicated cases. Future research on segmentation methods will be focused on the multiscale analysis of the proposed method and the topic of segmentation by parts. The proposed method is also applied to large image databases to examine its performance. In the performanceevaluation procedure, the detection rate is above 87.7% and the false alarm rate lies between 1.4% and 24% by human observation. The results are shown to be very promising for images with different lightings, backgrounds and occlusions. Future work on performance evaluation can focus on the selection (or design) of measures. Some results in the area of content based image retrieval may be useful for this topic. Techniques such as manifold learning may be applied for the postprocessing for color information, which may potentially improve the performance of the proposed method. A systematic evaluation of the proposed method will be performed on large image databases, and the results will be provided in future publications.

References

[1] S. Agarwal, A. Awan, and D. Roth, "Learning to detect objects in images via a sparse, part-based representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1475–1490, Nov. 2004.

[2] L. Ambrosio, V. Caselles, S. Masnou, and J. M. Morel, "Connected components of sets of finite perimeter and applications to image processing," *J. Eur. Math. Soc.*, vol. 3, no. 1, pp. 213–266, 2001.

[3] J. Aujol, G. Aubert, and L. Blanc-Féraud, "Wavelet-based level set evolution for classification of textured images," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1634–1641, Dec. 2003.

[4] C. Ballester and V. Caselles, "The M-components of level sets of continuous functions in WBV," *Pub. Matemátiques*, vol. 45, pp. 477–527, 2001.

[5] C. Ballester, V. Caselles, L. Igual, and L. Garrido, "Level lines selection with variational models for segmentation and encoding," *J. Math. Imag. Vis.*, vol. 27, no. 1, pp. 5–27, 2007.

[6] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, Jun. 1986.

[7] V. Caselles, "Topographic maps and local contrast changes in natural images," *Int. J. Comput. Vis.*, vol. 33, no. 1, pp. 5–27, 1999.

[8] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. J. Comput. Vis.*, vol. 22, no. 1, pp. 61–79, 1997.

[9] V. Caselles, J. L. Lisani, J. M. Morel, and G. Sapiro, "Shape preserving local histogram modification," *IEEE Trans. Image Process.*, vol. 8, no. 2, pp. 220–230, Feb. 1999.

[10] V. Caselles and P. Monasse, "Grain filters," *J. Math. Imag. Vis.*, vol. 17, no. 3, pp. 249–270, 2002.

[11] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.

[12] B. Coll and J. Froment, "Topographic maps of color images," in *Proc. Int. Conf. Pattern Recognition*, Barcelona, Spain, Sep. 2000, vol. 3, pp. 613–616.

[13] D. Cremers and G. Funka-Lea, "Dynamical statistical shape priors for level set based sequence segmentation," *Variational and Level Set Methods Comput. Vis.*, pp. 210–221, 2005.

[14] D. Cremers, F. Tischhauser, J. Weickert, and C. Schnorr, "Diffusion snakes: Introducing statistical shape knowledge into the Mumford–Shah functional," *Int. J. Comput. Vis.*, vol. 50, no. 3, pp. 295–313, 2002.
[15] Y. Chen *et al.*, "Using prior shapes in geometric active contours in a variational framework," *Int. J. Comput. Vis.*, vol. 50, no. 3, pp. 315–328, 2002.
[16] D. Freedman, R. J. Radke, and T. Zhang, "Model-based segmentation of medical imagery by matching distributions," *IEEE Trans. Med. Imag.*, vol. 24, no. 3, pp. 281–292, Mar. 2005.
[17] D. Freedman and T. Zhang, "Active contours for tracking distributions," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 518–526, Apr. 2004.
[18] R. C. Gonzalez and R. E. Woods*, Digital Image Processing*, 2nd ed. New York: Pearson Education, 2002.
[19] F. Guichard and J.-M. Morel, Image Analysis and P.D.E.s Preprint [Online]. Available: http://www.citeseer.ist.psu.edu/guichard01image.html, 2001
[20] L. Guigues, J. P. Cocquerez, and H. L. Men, "Scale-sets image analysis," *Int. J. Comput. Vis.*, vol. 68, no. 3, pp. 289–317, 2006.
[21] A. Hanbury and J. Serra, "Mathematical morphology in the HLS colour space," in *Proc. Brit. Machine Vision Conf.*, Manchester, U.K., Sep. 2001, pp. 451–460.
[22] A. Hanbury and J. Serra, A 3d-Polar Coordinate Colour Representation Suitable for Image Analysis, PRIP, T.U. Wien, Tech. Rep. PRIP-TR-77, 2002.
[23] M. Heiler and C. Schnorr, "Natural image statistics for natural image segmentation," *Int. J. Comput. Vis.*, vol. 63, no. 1, pp. 5–19, 2005.
[24] S. Jehan-Besson, M. Barlaud, and G. Aubert, " DREAM²S: Deformable regions driven by an Eulerian accurate minimization method for image and video segmentation," *Int. J. Comput. Vis.*, vol. 53, no. 1, pp. 45–70, 2003.
[25] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi, "Gradient flows and geometric active contour models," in *Proc. Int. Conf. Computer Vision*, Boston, MA, Jun. 1995, pp. 810–815.
[26] M. Leventon, O. Faugeraus, and W. Grimson, "Level set based segmentation with intensity and curvature priors," in *Proc. IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, Hilton Head Island, SC, Jun. 2000, pp. 4–11.
[27] M. E. Leventon, W. Eric, L. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours," in *Proc. Int. Conf. Computer Vision and Pattern Recognition*, Hilton Head Island, SC, Jun. 2000, pp. 316–323.
[28] J. L. Lisani, L. Moisan, P. Monasse, and J. M. Morel, "On the theory of planar shape," *Multiscale Model Simul.*, vol. 1, no. 1, pp. 1–24, 2003.
[29] P. Monasse, "Morphological representation of digital images and application to registration," Ph.D. dissertation, Ceremade, Paris, 2000.
[30] P. Monasse and F. Guichard, "Fast computation of a contrast-invariant image representation," *IEEE Trans. Image Process.*, vol. 9, no. 5, pp. 860–872, May 2000.
[31] D. Mumford and J. Shah, "Optimal approximation by piecewise smooth functions and associated variational problems," *Commun. Pure Appl. Math.*, vol. XLII, pp. 577–685, 1989.
[32] S. Osher and R. Fedkiw*, Level Set Methods and Dynamic Implicit Surfaces*. New York: Springer, 2003.
[33] S. Osher and J. Sethian, "Front propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations," *J. Comput. Phys.*, vol. 79, pp. 12–49, 1988.
[34] N. Paragios and R. Deriche, "Geodesic active contours for supervised texture segmentation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Ft. Collins, CO, Jun. 1999, pp. II:422–II:427.
[35] N. Paragios and R. Deriche, "Coupled geodesic active regions for image segmentation: A level set approach," in *Proc. Eur. Conf. Computer Vision*, Dublin, Ireland, Jun. 2000, vol. II, pp. 224–240.
[36] C. Samson, L. Blanc-Féraud, G. Aubert, and J. Zerubia, "A level set model for image classification," *Int. J. Comput. Vis.*, vol. 40, no. 3, pp. 187–197, 2000.
[37] J. Sethian*, Level Set Methods and Fast Marching Methods*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1999, Cambridge Monograph on Applied and Computational Mathematics.
[38] A. Tsai, A. Yezzi, and A. S. Willsky, "Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification," *IEEE Trans. Image Process.*, vol. 10, no. 8, pp. 1169–1186, Aug. 2001.
[39] D. Tschumperle, C++ Template Image Processing Library [Online]. Available: http://www.cimg.sourceforge.net/ 2004
[40] Z. Tu, X. Chen, A. L. Yuille, and S. C. Zhu, "Image parsing: Unifying segmentation, detection, and recognition," *Int. J. Comput. Vis.*, 2005, Marr Prize Issue.

Yongsheng Pan (S'03–M'08) received the B.S. degree in electrical engineering from Harbin Engineering University, China, in 1999, the M.S. degree in biomedical engineering from Shanghai Jiaotong University, China, in 2002, and the Ph.D. degree in electrical engineering from the University of Tennessee, Knoxville, 2007.

He is a Postdoctoral Researcher at the University of Utah, Salt Lake City. His research interests include image and video analysis using partial differential equations, mathematical morphology, particle filters, and parallel computation.

J. Douglas Birdwell (S'73–M'78–SM'85–F'99) received the B.S. and M.S. degrees in electrical engineering from the University of Tennessee (UT), Knoxville, and the Ph.D. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1978.

He is a Professor of electrical engineering and computer science at UT. He joined the faculty at UT in 1978 and is presently the Director of the Laboratory for Information Technologies, which develops secure distributed information systems and analysis tools. His research interests include high performance database design with applications in bioinformatics, parallel computation and local balancing, control systems, signal processing, and artificial intelligence.

Seddik M. Djouadi (S'91–M'99) received the B.S. degree (with first class honors) from the Ecole Nationale Polytechnique, Algiers, Algeria, the M.A.Sc. degree from the Ecole Polytechnique, Montreal, QC, Canada, and the Ph.D. degree from McGill University, Montreal, all in electrical engineering, in 1989, 1992, and 1999, respectively.

He is currently an Assistant Professor in the Electrical and Computer Engineering Department, University of Tennessee, Knoxville. He was an Assistant Professor in UALR, and held postdoctoral positions in the Air Force Research Laboratory and the Georgia Institute of Technology, where he was also with American Flywheel Systems, Inc. His research interests are modeling and power control for wireless networks, robust control, control under communication limitations, active vision, and identification.

Prof. Djouadi received the Ralph E. Powe Junior Faculty Enhancement Award 2005, four U.S. Air Force Summer Faculty Fellowships, the Tibbet Award with AFS, Inc., in 1999, and the ACC Best Student Paper Certificate (best five in competition) in 1998. He was selected by Automatica as an outstanding reviewer for 2003–2004.