Merging toward natural clusters

Zhi-gang Tan
Nelson H. C. Yung
The University of Hong Kong
Department of Electrical and Electronic
  Engineering
Pokfulam Road
Hong Kong
China
E-mail: zgtan@eee.hku.hk

Abstract. To findout how many clusters exist in a sample set is an old yet unsolved problem in unsupervised clustering. This problem inevitably occurs in region merging/growing, a well studied and popular technique in image segmentation. Region merging usually needs a stop criterion. The stop criterion is not automatically determined and often has to be set manually to arrive at a sensible segmentation, which is rather difficult for natural images. To address this problem, we present a robust stop criterion that is based on a novel distinctness predicate for adjacent regions. The predicate discerns distinct regions by examining the evidence of the boundary between neighboring regions. Requiring that every region should be distinct from each other, the proposed method is able to choose a stop point where a natural partition is most likely. Under a region merging framework, we demonstrate the effectiveness of the stop criterion using two merging criterion: one based on optimizing a global functional, and another based on a local criterion. Experimental results and comparison are given at the end. © 2009 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.3183892]

1 Introduction

Image segmentation, a technique that aims to divide the input image into nonoverlapping homogeneous regions, is a fundamental operation in many content-based applications of computer vision such as image query, image database organization, and object recognition. Image segmentation techniques are expected to reveal the underlying image structure, i.e., to find clusters that are natural in the image data. Clustering the data according to their features, either in a supervised or nonsupervised manner, is a natural solution. Methods such as region growing,[1,2] region split and merging, the hierarchical stepwise optimization algorithm (HSOA),[3] mean-shift transform,[4] and graph cutting[5–7] can all be tagged as unsupervised clustering methods.

A long standing problem in unsupervised clustering is to determine the number of clusters in the dataset. This problem is also known as the problem of validity and is largely unsolved.[8,9] Particularly in image segmentation, this problem is intrinsically challenging, since segmentation is an ill-posed problem and there is no comprehensive theory at the moment to evaluate what good segmentation is.

This work attempts to solve the problem of validity under the region merging framework. Though probably the oldest, region growing/merging as a simple and effective method is still widely used in many other modern segmentation techniques, either as a stand-alone method[10] or a postprocessing step. For example, region competition,[11] JSEG,[12] and edge flow[13] all use region merging to reduce oversegmentation.

In region growing/merging methods, finding the number of regions is equal to determining the stop criterion, which is closely related to the regularization parameter between data fidelity and model complexity.[14,15] With more regions, the output image is closer to the input image, while the model is more complex. A good balance of data fidelity and model complexity is usually hard to find[8] and often differs in applications. Due to such difficulty, many image segmentation methods that adopt region growing/merging techniques only leave the issue to the manual setting of the stop criterion, such as JSEG[12] and efficient graph segmentation.[10] These methods limit themselves in applications where more machine intelligence is needed. It should be noted that even manual setting is not an easy task. In the experiments presented later on, it is found that threshold(s) of the stop criterion are often related to the size of the image, size of the objects, homogeneity nature of regions, as well as noise level, which are unknown before the ground truth is known.

Unlike other aspects of image segmentation, the stop criterion problem is less addressed. Only a few references exist in the literature. For instance, hierarchical clustering methods, which merge the most similar regions at each step to form a hierarchical representation of the image, need to set a threshold in the cut level to attain a natural grouping. In hierarchical clustering methods, the usual assumption is that a significant disparity in the hierarchical levels indicates the presence of natural groupings.[8] Conceptually this makes sense, because the dissimilar regions begin to merge only until a certain point where all similar regions are merged already. This method works well for images with regions clearly separated in the feature space. However, in natural images, regions are often not compact or clearly separated in the feature space. Not surprisingly, the method only works for simple images.

This problem also has been modeled as searching for an

optimal regularization parameter in the Mumford-Shah model using classical L-curves.[14,15] The L-curve is a plot of solution size against approximation error. The optimal regularization parameter is the value corresponding to a convex corner in the L-curve. With this value, the stop criterion can be determined. It has been reported[14,15] that in image segmentation cases the L-curves often have no corners or have multiple corners, thus making it difficult to choose a correct segmentation. Experiments by the authors find that even when such a corner exists, segmentation may not be reasonable.

A recent technique by He et al.[16] presents a novel idea by minimizing the distance from an ideal assumed partition to a possible partition. The ideal solution is assumed to have zero intravariance and greatest intervariance. Under their definition of intravariance and intervariance, a partition with the smallest Mahalanobis distance is chosen as the optimal solution. This method works well for simple images. However, when the input is complex, the segmentation result becomes severely undersegmented.

We present a novel method for solving the problem. The main idea is that instead of solving the big problem of validity directly, we resolve a simpler question of how to discern two regions that are distinct, i.e., sufficiently different. If a distinctness predicate exists, all adjacent regions can be tested by it. A meaningful segmentation naturally requires that all adjacent regions are distinct from each other. Under this requirement, segmentation can be achieved with the number of regions determined automatically. We propose a distinctness predicate using a divide-and-conquer technique, which divides the regions into small atom segments (atom segments are defined as the initial segments obtained by oversegmentation, see Sec. 3 for details), and solves the distinctness of small atom segments by a statistical test. Then it determines the distinctness of the two regions by combining the distinctness of the atom segments. We demonstrate the effectiveness of the method in segmentation using two merging criteria: one based on optimizing a global functional and the other based on a local criterion.

The rest of the work is organized as follows. Section 2 introduces the merging framework and mathematical notations. Section 3 explains the oversegmentation step and the reasons why oversegmentation is important. Section 4 describes the details of the proposed distinctness predicate. Section 5 outlines the two algorithms based on the distinctness predicate: one based on a global criterion and the other based on making adaptive local decisions. Section 6 presents some experimental results and comparison with other methods. Section 7 concludes the work.

2 Merging Framework and Mathematical Notations

Many region merging methods start with an oversegmentation result, and then merge those regions based on a similarity measure such as region homogeneity or edge integrity.[17] The methods discussed in Refs. 18–21 all belong to this category. The merging framework adopted here also belongs to this category, which is summarized in the diagram shown in Fig. 1. In essence, the input image is first oversegmented to give M initial segments. Then a sequence of merging steps is applied iteratively on the initial seg-
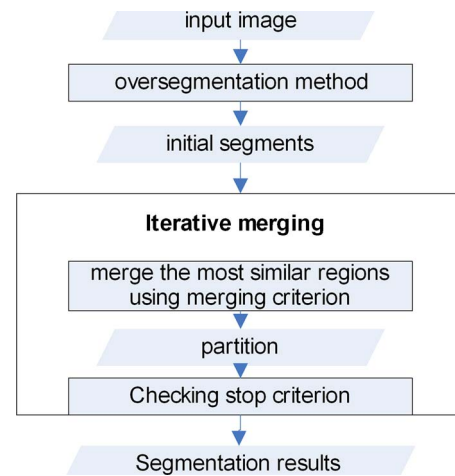

Fig. 1 Illustration of the merging framework (see text for details).

ments and generates a series of K-region partitions, $K = M, M-1\ldots$. At each merging step, two adjacent regions whose merging criterion value is the smallest are merged to form a new partition. If the new partition satisfies the stop criterion, the iterative merging process stops and the partition is generated as the output segmentation result.

Clearly, the two key questions in region merging are as follows.

1. How do we define the stop criterion, i.e., the condition when the merging stops? This question is equivalent to finding the number of regions.
2. How do we define the merging criterion or the similarity measure? The merging criterion is used to govern the order of merging. In each iteration, the two adjacent regions with the smallest value according to the merging criterion are chosen to be merged.

This work deals with both questions. Question 1 points toward a distinctness predicate that discerns whether two regions are different. The stop condition is that every region is distinct from each other. Question 2 points toward a merging criterion based on stepwise optimizing a global functional and also a modified criterion with local constraints.

For reader convenience, the mathematical notations are introduced here: $|\cdot|$ stands for cardinality and $\|\cdot\|$ stands for an L2 norm. The image I defined over $\Lambda \subset Z^2$, a portion of the 2-D integer grid, contains $|I|$ pixels, each being represented as an intensity value for gray images and red-green-blue values for color images. The boundaries of the regions are denoted as Γ.

3 Initial Oversegmentation

In our realization of the method, the initial segments are produced by a marker-controlled morphological watershed transform[16] of the gradient magnitude image. In principle, this transform finds small smooth patches in the topology of the magnitude map. Since the gradient inside each patch is small, the intensity/color inside the patch does not vary much. It should be noted that the initial oversegmentation is not limited by watershed algorithms as long as the bound-
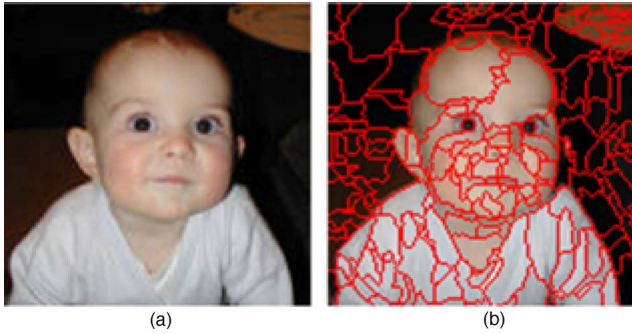
Fig. 2 Initial segments: (a) original image, and (b) initial segments with edges superimposed on original image.

Fig. 3 (a) Along the joint boundary of R_i and region R_j, six atom segments are shown. The five atom-segment pairs are {a, e}, {a, f}, {b, f}, {b, g}, and {c, g}. (b) Illustration of the atom-segment pair with supports S_1 and S_2, which are shaded.

ary information is sufficiently retained. Other methods such as quadrilaterals[22] and superpixels[23] can be used just as well. We choose the marker-controlled watershed transform for two reasons: 1. the markers are determined algorithmically and involve no human intervention; and 2. it gives comparable results while it is significantly faster for both gray and color images. An example is shown in Fig. 2. Interested readers can refer to He et al.[16] for details.

From a computational point of view, this oversegmentation step trims down the search space enormously. Before oversegmentation, every pixel is a possible edge pixel, while after oversegmentation, only those boundary pixels of the initial segments can be edge candidates. More precisely, the searching space is reduced from $2^{|I|}$ to $2^{|\Gamma|}$ ($|I|$ is the size of the input image and $|\Gamma|$ is the length of all boundaries Γ). In most cases, $|I| \gg |\Gamma|$. For the example in Fig. 2, a 130×132 image, we have $|I| = 17,160$, $|\Gamma| = 5130$, which means the latter searching space is only 2^{-12030} of the original! Other merits for analysis of region distinctness are given in Sec. 4.1. To reduce the ambiguity of terms, a segment obtained from the initial oversegmentation step is called an atom segment later on. Regions are formed by atom segments.

4 Distinctness Predicate for Regions

This section introduces the distinctness predicate for regions, which is either 1 (true) if the two regions are distinct, or 0 (false) if not. In this section, we first review the background and the divide-and-conquer idea of the proposed distinctness predicate for regions. Then we describe the details of the distinctness predicate for atom segments. In Sec. 4.3, we define the overall distinctness predicate for regions.

4.1 Overview of Proposed Distinctness Predicate

To determine two regions that are distinct, i.e., sufficiently different from each other, is not a trivial task. Conventionally, two clusters are thought to be distinct when they are distant enough in the feature space. In natural images, noise, shading effects, and uneven illumination are not uncommon. These lead to different parts of the same region having different feature values, i.e., making them distinct in the feature space, thus producing false boundaries. On the other hand, two truly distinct regions with low contrast between each other may not be discerned if a large threshold is set to reduce oversegmentation in highly varied regions.
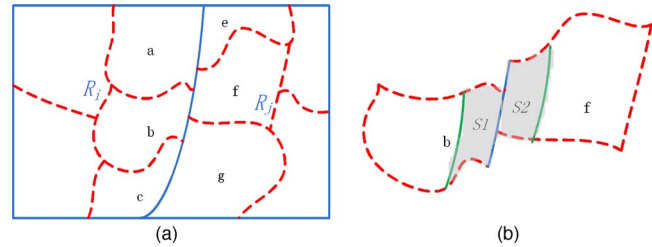
To cope with the aforementioned problem, we first divide the regions into less changeable atom segments by the oversegmentation method introduced in Sec. 3. The benefits of this division are threefold. First, this enables the subsequent analysis to adapt to the local context. Secondly, since variations inside atom segments are small, it is easier to analyze their distinctness from each other. For instance, simple statistical models such as Gaussian can fit the data inside atom segments well and thus reduces the complexity of distinctness analysis. [After oversegmentation, the size of the atom is usually (more than 95%) larger than 100 pixels, which is enough to ensure an accurate Gaussian model.] Third, the atom segments provide natural support for statistical estimation without concern about the shape of the sampling window.

After the division, we check the distinctness of the atom segments from their neighbors. Then we collect the results of the atom segments to decide whether regions are distinct from each other. For two neighboring regions, we check the distinctness of the small atoms along the common boundary. Similar to a voting system, the more distinctness evidence we collect on the small atom segments, the more likely the two regions are distinct.

An illustration is given in Fig. 3(a) to visualize the idea, where along the joint boundary of R_i and R_j, both have three atom segments (a, b, c for R_i and e, f, g for R_j) marked by red dashed lines. The joint boundary comprises the common edges of the neighboring atom segments pairs, which are marked by the solid blue line in Fig. 3(a).

To further enhance its adaptivity, we introduce a w-width band-shaped support for distinctness analysis. The distinctness of atom segment pairs is determined by feature distributions of the supports. Note that width w of the support is adaptive to its local content. The choice of the width is explained later in Sec. 4.2. Figure 3(b) shows the w-width supports S_1 and S_2 of a common edge between the pair {b, f}, marked by the shaded area where S_1 and S_2 are supports of atom segments b and f, respectively.

Using Fig. 3(a) as an example, to collect evidence of distinctness of atom segments, each pixel on the joint boundary has a vote. If the atom segment pair is distinct from each other, their common edge is labeled as "reliable." Every reliable edge votes to support the distinctness of the two regions (see Sec. 4.3). The more votes supporting the distinctness, the more likely the two regions are distinct. The idea of the distinctness predicate for regions is summarized as follows:
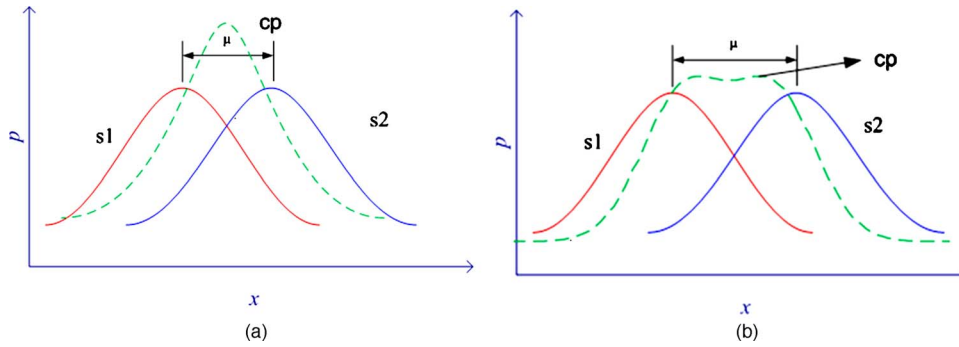
Fig. 4 Two cases of constructed distribution using samples drawing equally from S_1 and S_2. The solid lines are the original distributions of S_1 and S_2. The dashed line is the new distribution. s1 and s2 stand for the distribution of S_1 and S_2, respectively. cp stands for the constructed distribution in the statistical experiment. (a) s1 and s2 are not distinct. (b) s1 and s2 are distinct.

1. Along the joint boundary of two adjacent regions, all neighboring atom segments are grouped into atom segment pairs.
2. For every atom segment pair, a w-width band-shaped support is delineated along the common edge. If the atom pair is distinct, the common edge is labeled as reliable.
3. Finally, every reliable edge on the joint boundary votes to support the distinctness of the two regions. The more votes for distinctness, the more likely the two regions are distinct.

4.2 Distinctness Predicate for Atom Segments

The distinctness of an atom segment pair is confirmed if the following two conditions are satisfied.

1. The condition of different distributions: if two atom segments are distinct, then the supports of them have different feature distributions.
2. The condition of prominent edge: if two atom segments are distinct, then there should be a salient edge between them. This condition ensures that shading effects as seen in the sky or near a light source will not result in oversegmentation.

For the first condition, we fit a simple Gaussian model $N(\mu, \sigma^2)$ to the feature distribution of each side of the common edge and compare the distributions by their estimated parameters, which requires only $O(n)$ computational cost. This is much faster than the nonparametric measure earth mover distance (EMD) adopted in the compass operator,[24] which requires $O(n^3)$ computational complexity. Since the variation inside an atom segment is small and its feature distribution is tight and compact in the color space, it is reasonable to assume that each support follows a normal/Gaussian distribution. For the sake of simplicity, we only state our derivation for a single channel. For color images, two atom segments are found distinct if they are determined as distinct in at least one channel.

Let the feature distributions of S_1 and S_2 be denoted by s1 and s2. To compare whether the two distributions are sufficiently different, we conduct a statistical evaluation as follows. We generate a new distribution by drawing samples equally from S_1 and S_2. The histogram of the new distribution should have two peaks if s1 and s2 are sufficiently different from each other. In other words, the presence of two peaks in the new distribution indicates big differences between s1 and s2. On the contrary, if only one peak exists in the joint distribution, it is difficult to distinguish s1 and s2 in the joint distribution. Figure 4 depicts two cases of the new distribution. In Fig. 4(a), the difference of the two original distributions is small, and the new distribution has only one peak; whereas in Fig. 4(b), the difference of the two original distributions is large, and the new distribution depicts two peaks.

More formally, the atom segments are modeled as piecewise constants with identically independent distributed additive Gaussian noise $\eta \sim N(0, \sigma^2)$. For any pixel within the atom segment,

$$I = \mu_o + \eta, \tag{1}$$

such that I follows a normal distribution $N(\mu_o, \sigma^2)$. For S_1 and S_2, the estimated Gaussian distributions are denoted as $s1 \sim N(\mu_1, \sigma^2)$, $s2 \sim N(\mu_2, \sigma^2)$.

The new distribution, which is constructed by drawing samples equally from s1 and s2, can be represented as a mixture density using the estimated parameters shown as:

$$f_c(x|\mu_1, \mu_2, \sigma) = \frac{1}{2} f(x|\mu_1, \sigma) + \frac{1}{2} f(x|\mu_2, \sigma). \tag{2}$$

If the mixture density function is bimodal, then it has a local minimum at $x_0 = 1/2(\mu_1 + \mu_2)$, thus the Eqs. (3) and (4) should be satisfied.

$$f_c'(x_0) = 0, \tag{3}$$

$$f_c''(x_0) > 0. \tag{4}$$

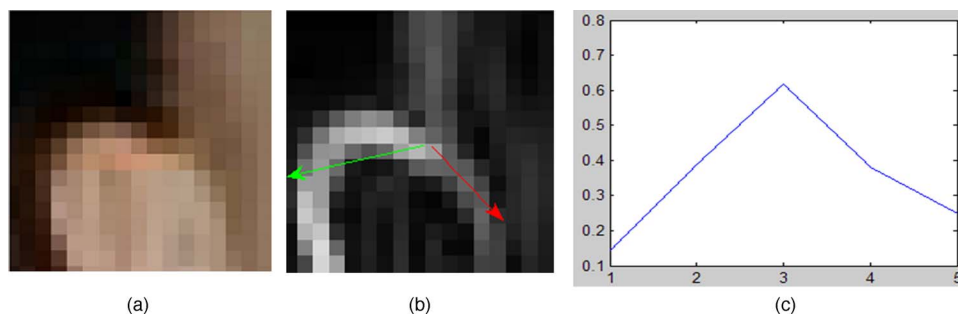It is easy to verify Eq. (3). By expanding the left-hand side of Eq. (4), we obtain

Fig. 5 (a) Original image. (b) Gradient magnitude map. Green and red arrows indicate the gradient vector and normal direction, respectively. (c) Gradient magnitude values along the normal direction. (Color online only.)

$$f_c''(x_0) = \frac{1}{2\sigma^3\sqrt{2\pi}}\left\{-\exp\left[-\frac{(x-\mu_1)^2}{2\sigma^2}\right]\right.$$

$$-\exp\left[-\frac{(x-\mu_2)^2}{2\sigma^2}\right]\right\}$$

$$+\frac{1}{2\sigma^5\sqrt{2\pi}}\left\{(x-\mu_1)^2\exp\left[-\frac{(x-\mu_1)^2}{2\sigma^2}\right]\right.$$

$$\left.+(x-\mu_2)^2\exp\left[-\frac{(x-\mu_2)^2}{2\sigma^2}\right]\right\}. \tag{5}$$

By substituting $x_0 = 1/2(\mu_1 + \mu_2)$ and Eq. (5) into Eq. (4), we obtain

$$f_c''\left(\frac{\mu_1+\mu_2}{2}\right) = \frac{1}{4\sigma^5\sqrt{2\pi}}$$

$$-\exp\left[-\frac{(\mu_2-\mu_1)^2}{8\sigma^2}\right][(\mu_2-\mu_1)^2-4\sigma^2]$$

$$> 0. \tag{6}$$

Equation (6) results in the inequality as given in Eq. (7),

$$(\mu_2 - \mu_1)^2 > 4\sigma^2. \tag{7}$$

As such, the condition of different distributions can be expressed as $(\mu_2 - \mu_1)^2/\sigma^2 > 4$. Since $(\mu_2 - \mu_1)^2/\sigma^2$ is the signal-noise ratio (SNR) for the estimated difference, essentially this condition requires the SNR to be large for determining the distinctness. The computation overhead is approximately $O(n)$, linear in terms of the number of pixels for estimating μ_1, μ_2, and σ^2. In addition, because μ_1, μ_2, and σ are all locally estimated, the distinctness predicate has the adaptive nature to local context. Note that width w of the support is variable for each common edge. The width that corresponds to the largest SNR estimated by varying w from 2 to 16 is set as the optimal width. This is similar to finding a maximum in the small scale space.

For the condition of prominent edge, we check whether the gradient magnitude value at each pixel on the common edge is a local maximum along its normal direction. This is similar in spirit with nonmaxima suppression in Canny's detector, but differs in the respect that Canny uses direction of gradient as the suppression direction. It has been pointed out that the direction of the gradient is not always perpendicular to the edge direction,[24] which can cause problems at junctions and corners. Under Canny's assumption, a function has zero derivatives along the edge direction. In practice, this derivative component is often nonzero, i.e., the gradient is not normal to the edge. For junctions where two edges intersect, one edge has to be suppressed, since nonmaxima suppression only keeps the maxima along the gradient direction. The proposed method is free from this problem. Figure 5 depicts an illustration of how the prominent edge condition is checked. Figures 5(a) and 5(b) are the input image and the view of gradient magnitude, respectively. Figure 5(b) shows the normal direction of the common edge and the direction of the gradient vector at the center of Fig. 5(a). Figure 5(c) plots the gradient magnitude values along the normal direction in the support.

4.3 Reliable Edge between Atom Segments

The distinctness predicate in essence examines whether there is evidence that a boundary exists between a pair of atom segments. If two atom segments are distinct, their common edge is then marked as "reliable." Some examples of reliable edges founded by evaluating distinctness of the atom segments are depicted in Fig. 6. Figure 6(a) is a synthesized image with gradual changing illumination from Ref. 25. Figure 6(b) is also a synthesized image but corrupted by Gaussian noise. Figure 6(c) is a natural "woman" image. A comparison is made with the Canny edge detector.[26] Using the default hysteresis threshold setting, Canny's detector generates many false edges compared with the reliable edges found by the distinctness predicate. Moreover, note that edges around the junction points of the woman image, highlighted by the red circles, are not connected due to Canny's nonmaxima suppression, while the reliable edges consistently produce connected junctions.

It is true that the false edges detected by the Canny detector can be improved by manually fine tuning the parameters. However, Canny's edge detector lacks an adaptive nature. In this sense, the proposed method is superior. Canny's hysteresis thresholding uses a low threshold to pick up edge candidates and a high threshold to identify strong seeds for edges, aiming to eliminate trivial features and improve overall performance. In practice, it is common that some real edges are higher than the lower threshold but are not connected to any strong edges. (An excellent example is given in He and Yung's work[27] to illustrate this
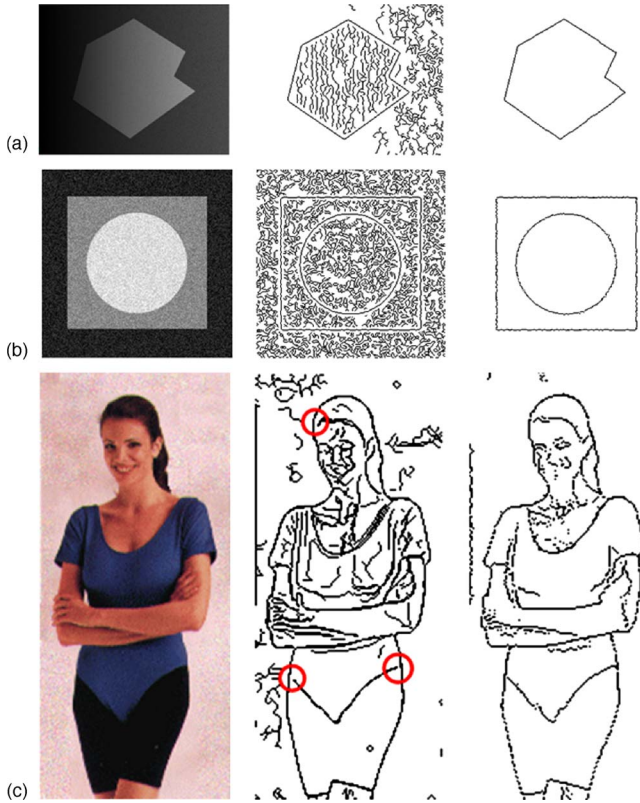
Fig. 6 Comparison of reliable edge map and Canny edge map. First column, original images: (a) nonuniform case, (b) noisy case, and (c) natural image. Second column, Canny edge map. Third column, reliable edge maps by the proposed method.

point). Under such circumstances, Canny's edge detector fails even when fine tuning the parameters, while the proposed method still works well due to its adaptivity.

In summary, the distinctness predicate for atom segments compares distributions in local support of the two neighboring atom segments, which provides the method adaptivity to a local context. Furthermore, unlike many edge operators that only use fixed-shape sampling windows, the distinctness predicate extends the sampling window to arbitrary shapes formed by the watershed transform. Since the common edge of atom segments is labeled as reliable as long as they are distinct, the distinctness predicate for atom segments is equivalent to a reliable edge labeling method.

4.4 Distinctness Predicate for Regions

To combine evidence of distinctness of atom segments, the distinctness predicate for regions is designed as a voting mechanism. Let Γ_{ij} be the boundary between R_i and R_j (i.e., $\Gamma_{ij} = \partial R_i \cap \partial R_j$) and E denote the set of common edges labeled as reliable on Γ_{ij}, then the distinctness predicate for R_i and R_j is defined as

$$l(R_i, R_j) = \begin{cases} 1, & g(\Gamma_{ij}) \geq T_0 \\ 0, & g(\Gamma_{ij}) < T_0 \end{cases}, \tag{8}$$

where $g(\Gamma_{ij}) = \Sigma_{l \in E} |l| / |\Gamma_{ij}|$. The threshold T_0 is related to the success rate of reliable edge detection on real bound-

aries. Section 6 has more discussions on how to set T_0.

5 Merging Criteria

In merging techniques there are two strategies: a global criterion and a local criterion. The global criterion is usually a carefully designed function/functional such as the Mumford-Shah functional,[28] the sum of squared error, and the minimal description length criterion,[11] where segmentation can be obtained by optimizing the criterion. One can also derive a merging criterion using a greedy optimization method. For example, the hierarchical stepwise optimization algorithm[3] derives the merging criterion by minimizing the sum of squared error criterion in a stepwise manner. The merit of having a global criterion is that the segmentation results are usually stable to local noise, as long as the global functional can reasonably capture the structure and characteristics of the input image. However, the result can be absurd when the input image deviates from the global model.

Unlike the global method, the local decision method only utilizes local information; for example, Brook, Kimmel, and Sochen[29] used local statistics to prevent heterogeneous regions from being merged. The local method provides a flexible alternative for keeping important local information of the input when the global criterion fails to capture its structure. However, this method is sensitive to noise.

We present two criteria. Criterion 1 (C1) is derived by minimizing a functional in a stepwise manner. Criterion 2 (C2) is a combination of C1 and a local constraint using the distinctness predicate for regions.

5.1 Criterion 1: Merging toward a Global Functional

Let $I : \Lambda \to \mathfrak{R}, \Lambda \subset Z^2$ represent an image. The objective is to find a partition $R = \{R_1, \cdots, R_K | \Lambda = \cup_i^K R_i, R_i \cap R_j = \varnothing, \forall i \neq j\}$ that optimizes the following functional,

$$E[f, \Gamma | K] = \alpha \sum_{i=1}^{K} \sum_{p \in R_i} [f_i(p) - I(p)]^2$$
$$- \lambda \sum_{\Gamma_{ij} \in \Gamma} \sum_{p \in \Gamma_{ij}} \frac{1}{|\Gamma_{ij}|} \|\nabla I(p)\|, \tag{9}$$

where p denotes a position on Λ and $\Gamma = \{\Gamma | \Gamma_{ij} = \partial R_i \cap \partial R_j \neq \varnothing, R_i, R_j \in R\}$ denotes the segmentation boundaries of the entire image with Γ_{ij} being any joint boundary between regions R_i and R_j. The two terms in Eq. (9) can be rewritten as

$$E_a[f, \Gamma | K] = \alpha \sum_{i=1}^{K} \sum_{p \in R_i} [f_i(p) - I(p)]^2, \tag{10a}$$

$$E_b[\Gamma] = -\lambda \sum_{\Gamma_m \in \Gamma} \sum_{p \in \Gamma_m} \frac{1}{|\Gamma_m|} \|\nabla I(p)\|. \tag{10b}$$

E_a encourages the output image to have small differences from the original image, and E_b encourages the region boundaries to be close to pixels with large gradient magnitude. α and λ control E_a and E_b's weight. Taking a stepwise
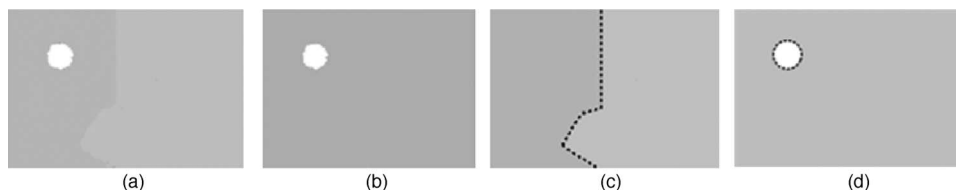
Fig. 7 A failure case for the global criterion (criterion 1) (with $K=2$). (a) Original image. (b) Ground truth. (c) Segmentation using C1 with $\alpha=1$ and $\lambda=0$. (d) Segmentation using C2 with $\alpha=1$ and $\lambda=0$. Region boundaries are denoted by dashed lines.

optimization of Eq. (9), the merging criterion is shown in Eq. (11) (see Appendix).

$$\delta(R_i, R_j) = \alpha \frac{|R_i| \cdot |R_j|}{|R_i| + |R_i|}(f_i - f_j)^2 + \lambda \frac{1}{|\Gamma_{ij}|} \sum_{p \in \Gamma_{ij}} \|\nabla I(p)\|. \quad (11)$$

Stepwise optimization is a greedy method and may be trapped by local minima. However, in image segmentation, the results are pleasant and acceptable to human eyes.[18,28,30] A theoretical analysis by Beaulieu and Goldberg[3] also shows that oversegmentation error can be corrected in an iterative merging process, while undersegmentation error is kept low.

5.2 Criterion 2: Merging Using Local Constraints

One limitation of criterion 1 is that the parameters α and λ control the merging order. If they are incorrectly chosen, the final result may not be appealing. Since the first term is indeed a sum of squared error, minimizing it cannot reveal the underlying structure of the input data when there are great differences in the number of samples/pixels in different clusters/regions.[8] The global criterion has a bias to merge small regions. A failure case is depicted in Fig. 7.

On the other hand, for a merging algorithm, different parameter values would change the merging order of regions. The reason why the final segmentation result has errors is because some merges are wrong, i.e., heterogeneous regions are merged. The idea of criterion 2 is to prevent distinct regions from being merged using the distinctness predicate developed in Sec. 4. With this safeguard, the merging method becomes more robust to different image contents, even when different parameters are used. Criterion 2 is given in Eq. (12),

$$\delta(R_i, R_j) = \alpha \frac{|R_i| \cdot |R_j|}{|R_i| + |R_i|}(f_i - f_j)^2 + \lambda \frac{1}{|\Gamma_{ij}|} \sum_{p \in \Gamma_{ij}} \|\nabla I(p)\|$$
$$+ D \cdot l(R_i, R_j). \quad (12)$$

Whenever R_i and R_j are distinct from each other, i.e., $l(R_i, R_j) = 1$, $\delta(R_i, R_j) \approx D$, where $D = \infty$, thus they will not be merged. Using C2, the problem as shown in Fig. 7 is eradicated.

6 Experimental Results, Analysis, and Comparison

In this section, experimental results are shown and compared with other methods. First, we introduce how to set the distinctness predicate and compare the results obtained by the proposed method with other stop criteria. Secondly,

we analyze the results obtained by setting different parameters for C1 and C2 and their advantages and disadvantages are summarized Finally, the proposed method is compared with a popular graph cut method.

6.1 Setting the Distinctness Predicate

The only parameter in the distinctness predicate $1(R_i, R_j)$ is the threshold T_0. Recalling that $g(\Gamma_{ij})$ in Eq. (8) is the ratio of reliable edges detected on the joint boundary, it is natural to require T_0 to be larger than 0.5 in a voting system. A small value of T_0 requires little distinctness and leads to a large number of regions, and a large value leads to a small number of regions. A good choice of T_0 should be a good balance between data fidelity and model complexity. Since the purpose of segmentation is to find boundaries of true regions, the partition that preserves most reliable edges and least unreliable edges can be considered a good balance. Therefore, T_0 is determined by the following equation,

$$T_0 = \arg \max_{T_0} \{|E_r| - |E_f| | 1 \geqslant T_0 \geqslant 0.5\}, \quad (13)$$

where E_r is the set of reliable edges detected on region boundaries, and E_f is the set of unreliable edges. Typically, T_0 determined by Eq. (13) ranges from 0.5 to 0.8.

6.1.1 Comparison with other stop criteria

In this section, we compare the proposed stop criterion based on the distinctness predicate with four other stop cri-
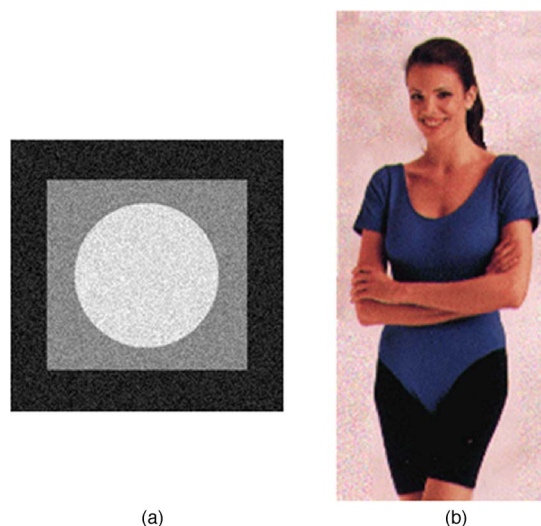

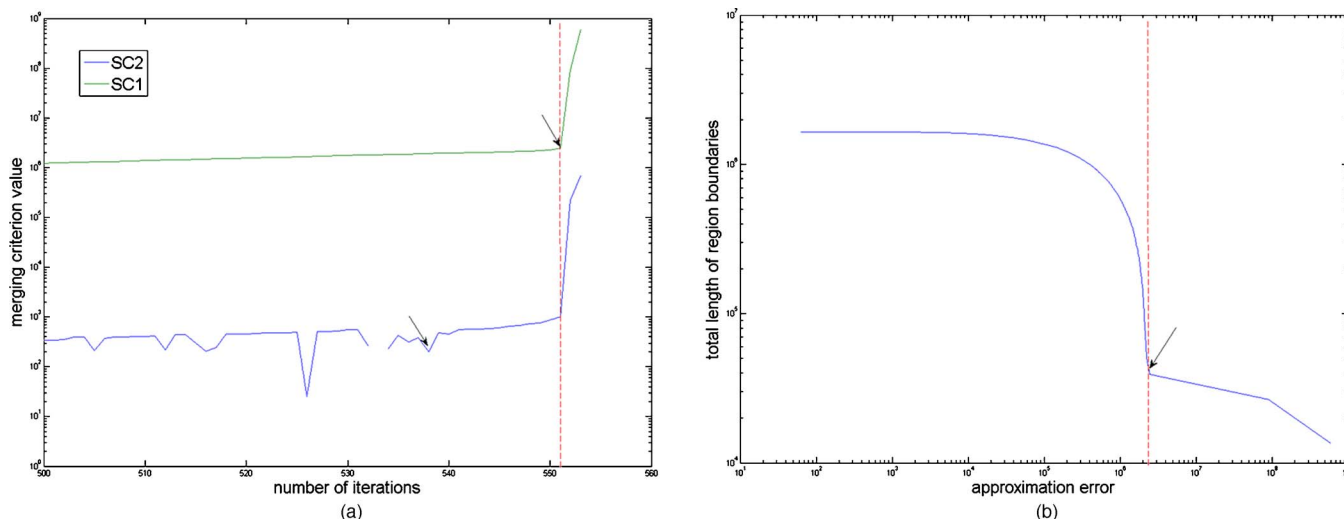**Fig. 8** Input images: (a) synthesized images and (b) woman image.

Fig. 9 Stop criterion applied to Fig. 9(a). (a) Plots of SC1 and SC2 at the last 50 iterations; and (b) plot of SC3. Red dashed line shows the final stop point. Arrows denote the stop point found by the corresponding stop criteria. (Color online only.)

teria. For convenience, the merging criterion used in the proposed method is C1 with $\alpha=1$ and $\lambda=0$. The five stop criteria are listed next.

1. Stop criterion 1 (SC1)—significant jump in hierarchical clustering:[8] it is a sudden change in the adopted criterion value. Graphically, the jump is a corner indicating a trend changing in the plot of merging criterion value against iteration.
2. Stop criterion 2 (SC2)—stop rule suggested in full lambda-schedule segmentation:[14] this stop point is the last significant local minimum in the plot of merging criterion values against the number of iterations. The merging criterion adopted in this method is $|R_i|\cdot|R_j|(f_i-f_j)^2/\{(|R_i|+|R_i|)|\Gamma_{ij}|\}$.
3. Stop criterion 3 (SC3)—L-curves suggested by Hansen and O'Leary:[15] it is a plot of the approximation error against the solution size. In this context, the approximation error is E_a of the first part of Eq. (10)

and the solution size is $|\Gamma|$. This method chooses a convex corner in the L-curve as the stop point.

4. Stop criterion 4 (SC4)—intervariance and intravariance metric by He et al.:[16] this method defines the optimal segmentation as a segmentation closest to an ideal solution that has zero intravariance and greatest intervariance. The distance metric used is Mahalanobis distance. Intravariance is defined as the approximation error as in the first part of Eq. (10), and intervariance is defined as the minimal value of the merging criterion of all adjacent region pairs.
5. The proposed method—distinctness predicate is applied to all adjacent region pairs. If the predicate is true for all region pairs, the iteration stops.

Note that SC1, SC2, and SC3 all rely on the shape of the curves. SC1 and SC2 use plots of values of merging criterion against iterations to determine the stop point, while SC3 uses a plot of solution size against approximation er-

Synthesized image					
Stop criterion	SC1	SC2	SC3	SC4	Proposed
Stop criterion threshold	2.451×10^6	1.009×10^3	3.911×10^4	Null	[0.50 1.00]
Number of regions	3	16	3	3	3
Segmentation results					

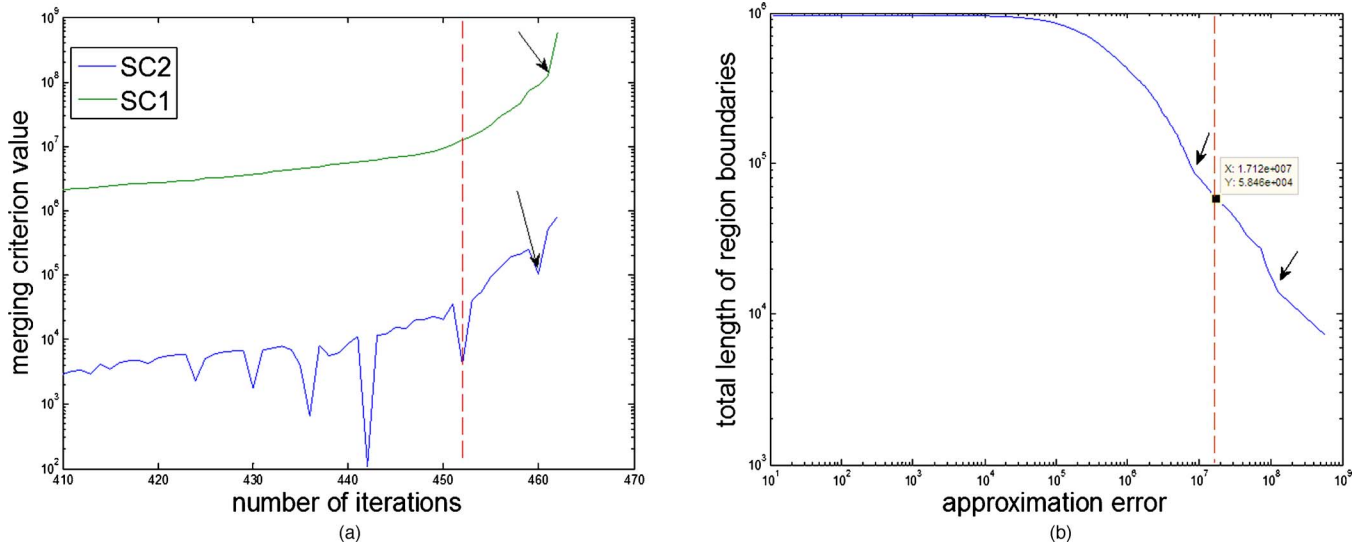**Fig. 10** Summary of different stop criteria on the synthesized image.

Fig. 11 Stop criterion applied to Fig. 9(b). (a) Plot of SC1 and SC2 at last 50 iterations. (b) Plot of L-curve.

ror. In contrast, SC4 and our proposed merging criterion use algorithmic means to determine the stop point. These stop criteria are applied to images in Fig. 8 and their graphs are compared. SC1 and SC2's curves are plotted together in one graph, and SC3's is plotted on a separate graph alongside the first one. SC4 and the proposed method are both summarized in a table.

These stop criteria are first applied to the synthesized image in Fig. 8(a). The curves of SC1, SC2, and SC3 are plotted in Fig. 9. There is a conspicuous corner in each

curve of SC1 and SC3, identified as stop points by them. These stop points correspond to a segmentation of three regions, which agrees with the ground truth. On the contrary, the stop point found by SC2 leads to a 16-region segmentation that differs substantially from the ground truth. SC1, SC3, SC4, and the proposed method all achieve the correct segmentation in this case. The output of these stop criteria are summarized in Fig. 10.

For the woman image in Fig. 8(b), Fig. 11 shows curves of SC1, SC2, and SC3. Unlike Fig. 9, these curves appear

Woman image					
Stop criterion	SC1	SC2	SC3	SC4	Proposed
Stop criterion threshold	1.974×10^6	7.815×10^6	1.393×10^5 Or 0.822×10^5	Null	[0.50 1.00]
Number of regions	3	4	3 or 15	6	11
Segmentation results					

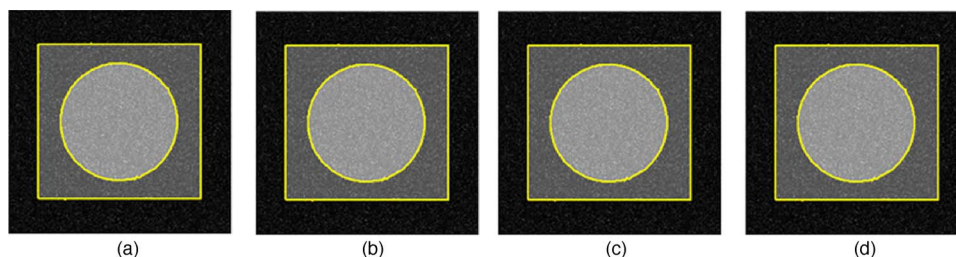**Fig. 12** Summary of different stop criteria on the "woman" image.

Fig. 13 Example 1 using C1 with different parameter settings: (a) $\alpha=1$, $\lambda=0$. (b) $\alpha=0.7$, $\lambda=0.3$. (c) $\alpha=0.4$, $\lambda=0$. (d) $\alpha=0.1$, $\lambda=0.9$.

smoother, which makes it difficult to identify any clear corners. The best choice for a corner is the point with the largest curvature, denoted by the arrows in the graph. SC1 has an obscure corner and SC3 has two such corners, which are undesirable for a stop criterion. In fact, the results obtained by SC1, SC2, and SC4 are apparently undersegmented. They all missed the boundaries between the background and legs, and between the hair and the face. SC3 has two possible corners that produce a 15-region partition or three-region partition; the 15-region one is obviously oversegmented, and the three-region one is undersegmented, as shown in Fig. 12. On the contrary, the result by the proposed stop criterion has 11 regions with all the important region boundaries correctly identified. Though the boundaries on the chest and left arm may be arguable, there is noticeable difference between the regions in these parts.

We have evaluated the proposed method over many different images, all of which show that our method outperforms other stop criterion. Besides better performance, a further merit of the proposed method is that T_0 is quite stable from one image to another, even when SC1, SC2, and SC3 change significantly. For instance, the threshold of SC1 for Fig. 9(a) is 2.451×10^6, while the threshold for Fig. 9(b) is 1.974×10^6. For the experiments conducted in Sec. 6.2, the same $T_0=0.6$ is used, which proves the performance of the proposed distinctness predicate.

6.2 Setting the Merging Criteria

For the merging criterion, the parameter α stresses small error between the output image and input image, and λ emphasizes gradient information. Usually E_a is a lot larger than E_b. To make them comparable, we normalize them before parameter setting.

6.2.1 Results of applying criterion 1 and corresponding analysis

Figure 13 depicts the results on a synthesized image that

contains regions of constant intensity corrupted by Gaussian noises with different strength. The results are not affected by different parameter settings and are robust to noise. This is expected because the synthesized image follows exactly the adopted mathematical model whose output image is piecewise constant and contaminated by additive Gaussian noise. Under such circumstances, the two terms E_a and E_b are not in conflict with each other, i.e., to minimize E_a is to minimize E_b, such that changing α and λ would not change the segmentation result.

However, there are circumstances when E_a conflicts with E_b, for example, images with nonuniform illumination or shading effects. Under such circumstance, minimizing E_b (maximizing gradient magnitude on boundary) may result in increasing E_a. α and λ in this case need to be carefully chosen, as depicted in Fig. 14.

More experimental results obtained by different parameter settings are shown in Fig. 15. These input images are natural images. We observe that segmentation results by C1 tend to be stable and are good at boundaries of large regions, but may lose small distinctive regions such as the eyes of the woman in the first row of Fig. 15 and the eyes on the baby face in the second row. This effect can be reduced by using criterion 2, where local constraint is combined.

6.2.2 Result of applying criterion 2 and corresponding analysis

As stated by Sec. 5.2, small distinctive regions are very likely missed by using criterion 1. Criterion 2 is designed to avoid such errors. Thus, it should have the ability to retain these small distinctive regions. Applying criterion 2 to the input images in Fig. 13 renders the exact same results, which are omitted here. For noneven illumination cases, it renders consistent results as depicted in Fig. 16 using different parameters. It is because incorrect merges can be
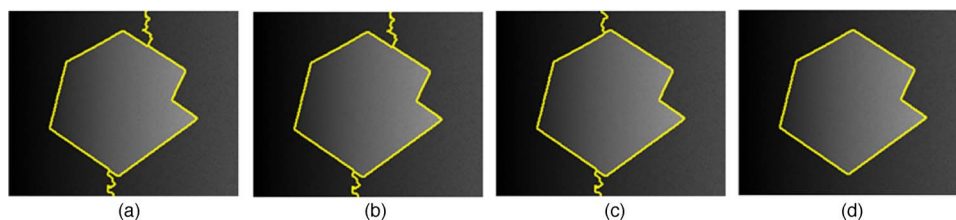

Fig. 14 Example 2 using C1 with different parameter settings: (a) $\alpha=1$, $\lambda=0$. (b) $\alpha=0.7$, $\lambda=0.3$. (c) $\alpha=0.4$, $\lambda=0.6$. (d) $\alpha=0.001$, $\lambda=0.999$.
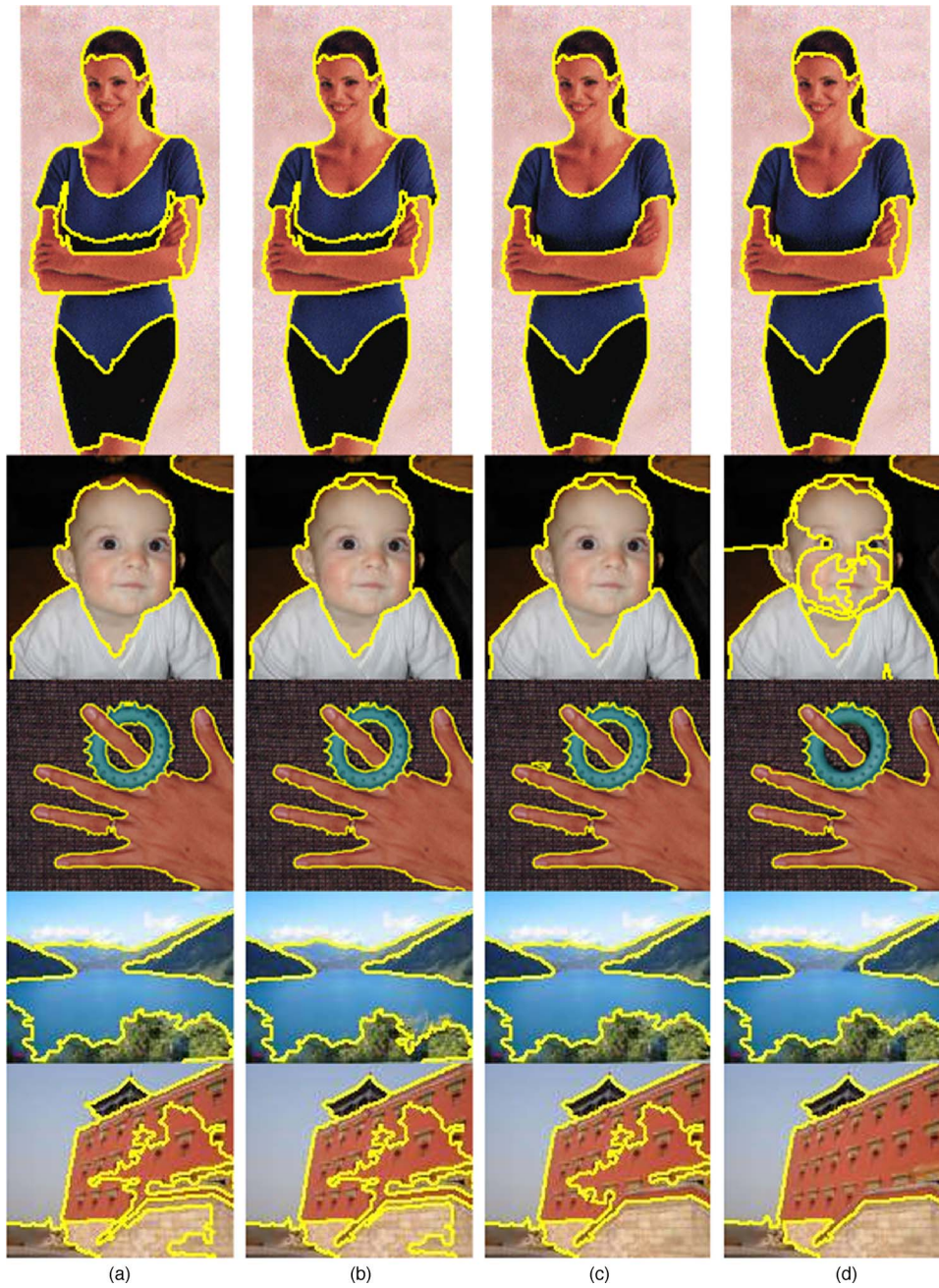
Fig. 15 Examples with different parameter settings: (a) $\alpha=1$, $\lambda=0$. (b) $\alpha=0.7$, $\lambda=0.3$. (c) $\alpha=0.4$, $\lambda=0.6$. (d) $\alpha=0.1$, $\lambda=0.9$.
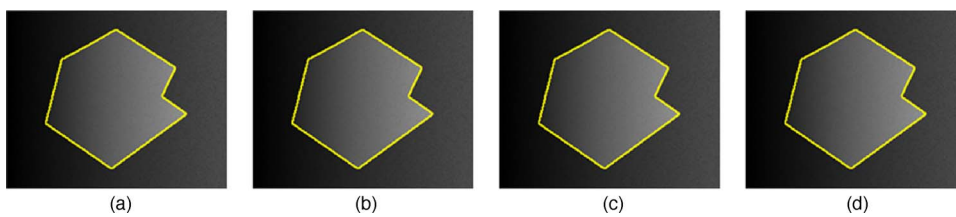

Fig. 16 Examples using C1 with different parameter settings: (a) $\alpha=1$, $\lambda=0$. (b) $\alpha=0.7$, $\lambda=0.3$. (c) $\alpha=0.4$, $\lambda=0.6$. (d) $\alpha=0.001$, $\lambda=0.999$.
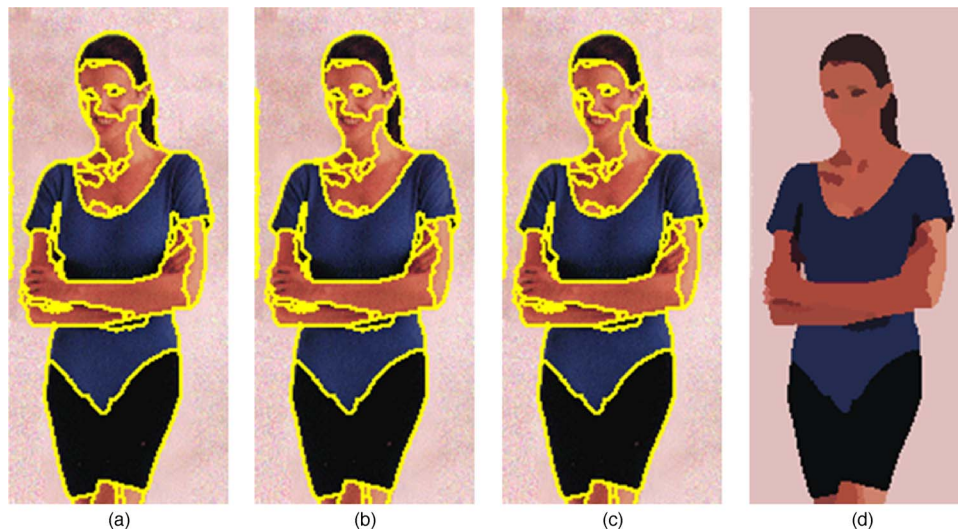
Fig. 17 Examples with different parameter settings: (a) $\alpha=1$, $\lambda=0$. (b) $\alpha=0.7$, $\lambda=0.3$. (c) $\alpha=0.4$, $\lambda=0.6$. (d) output image.

prevented, i.e., two distinct regions are not merged, and the merging method becomes more robust to different weight settings.

Figure 17 shows another example using different parameters. Note that the segmentation results are quite similar even using different parameters. Figure 17(d) depicts the output image. Compared with the results of criterion 1, the woman's eyes are retained as well as other small regions that are distinct from their neighbors. The dark yellow region with relatively low contrast on the left of the baby is detected, while it is missed in the results of criterion 1 at the cost of more false edges.

More natural images are tested in Fig. 18 using Criterion 2. These results show that more distinct details can be retained using criterion 2. However, we need to be aware that some small details, though distinct to their neighbors, may not contain important information and may not be significant in the final segmentation results.

6.3 Comparison with Normalized Cut

A brief comparison between the proposed method and normalized cut—an often cited graph cut method[6]—is depicted in Fig. 19. Since the normalized cut method needs to specify a region number, we set it to be the same as the number found by the proposed method for comparison sake. As can be seen in Fig. 19, the graph cut method fails to detect the object. For the "baby" and "woman" images, the normalized cut generates a lot more false edges than the proposed method using the two merging criteria. This is due to the fact that a normalized cut uses a dissimilarity measure constrained by spatial proximity, such that spatially distant pixels tend to have large dissimilarity and cannot be merged.

7 Conclusion and Future Work

The major contribution made in this work is that we propose a distinctness predicate to solve the problem of validity in image segmentation under the framework of region merging. Not only is the stop criterion based on the proposed distinctness predicate shown to perform better than many other stop criteria, but it also exhibits stability over different thresholds. In addition, we also show the distinct-
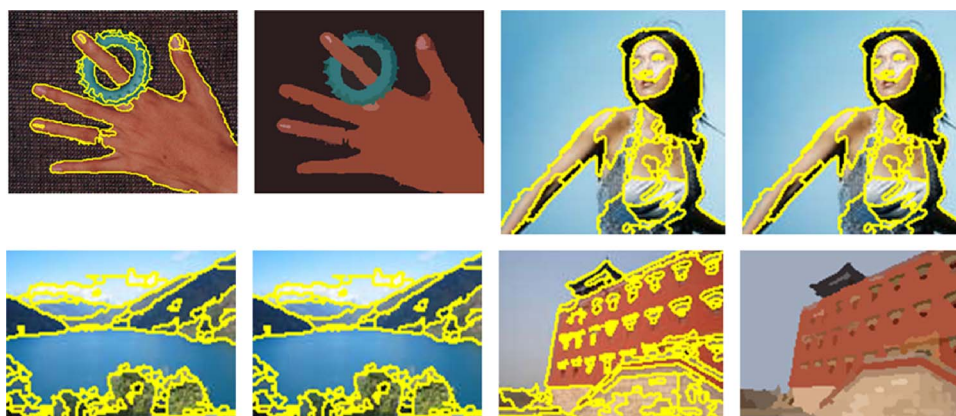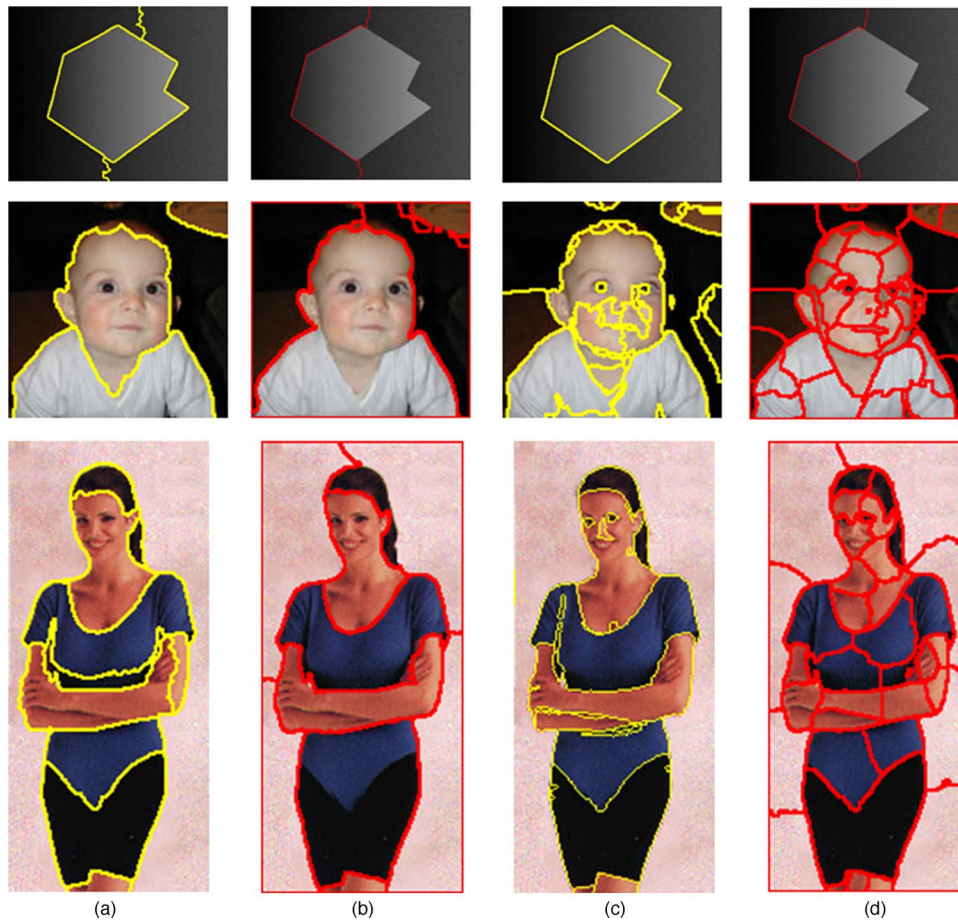

Fig. 18 More results.

Fig. 19 Results comparison with graph cut. (a) Proposed—C1. (b) Graph cut using segment number. from C1. (c) Proposed—C2. (d) Graph cut using segment number from C2.

ness measure as capable of detecting edges in color images and having significant advantage over the Canny edge detector in certain circumstances.

Another contribution of this work is that two new merging criteria are proposed: one is derived toward stepwise optimization of a global functional, shown to be stable to noise and local variations; another is derived by taking local constraints into account, which has an interesting property of being able to preserve small but distinct regions containing important information. Experiment results show that both criteria are able to arrive at a natural grouping.

As this distinctness predicate only considers color and intensity at the moment, the same idea could be extended to other features such as texture, which will be investigated in the future.

Appendix: Derivation of the Merging Criterion

From Eqs. (9) and (10), using the assumption that f is a piecewise constant, it is straightforward to show that $E[f, \Gamma | K]$ is minimized when $f_i = 1/|R_i| \iint_{R_i} I dx dy$, such that $E[f, \Gamma | K]$ is reduced to $E[\Gamma | K]$. This indicates that the value of Eq. (9) is only determined by the segmentation boundaries Γ, i.e., a partition. Let R_i and R_j be the concerned regions during merging. The value of E_a only changes over R_i and R_j [see Eq. (15)] as the value over other regions remains; and merging the two regions is equal

to removing the joint boundary between R_i and R_j. So the change of the second term E_b can be easily derived, which is shown in Eq. (16). The change of the energy function by merging R_i and R_j is

$$\delta(R_i, R_j) = \Delta E_a + \Delta E_b, \tag{14}$$

where ΔE_a and ΔE_b are respectively defined as:

$$\Delta E_a = \alpha \frac{|R_i| \cdot |R_j|}{|R_i| + |R_i|} (f_i - f_j)^2, \tag{15}$$

$$\Delta E_b = \lambda \frac{1}{|\Gamma_m|} \sum_{p \in \Gamma_m} \|\nabla I(p)\|. \tag{16}$$

References

1. R. Adams and L. Bischof, "Seeded region growing," *IEEE Trans. Pattern Anal. Mach. Intell.* **16**, 641–647 (1994).
2. H. P. Fan, G. H. Zeng, M. Body, and M. S. Hacid, "Seeded region growing: an extensive and comparative study," *Pattern Recogn. Lett.* **26**, 1139–1156 (2005).
3. J. M. Beaulieu and M. Goldberg, "Hierarchy in picture segmentation—a stepwise optimization approach," *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 150–163 (1989).
4. D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 603–619 (2002).

5. Z. Wu and R. Leahy, "An optimal graph-theoretic approach to data clustering—theory and its application to image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 1101–1113 (1993).
6. J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 888–905 (2000).
7. Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**, 1222–1239 (2001).
8. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed., John Wiley and Sons, Inc., Hoboken, NJ (2003).
9. D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, Prentice Hall, Englewood Cliffs, NJ (2003).
10. P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.* **59**, 167–181 (2004).
11. S. C. Zhu and A. Yuille, "Region competition: unifying snakes region growing and Bayes/MDL for multiband image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **18**, 884–900 (1996).
12. Y. N. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**, 800–810 (2001).
13. W. Y. Ma and B. S. Manjunath, "EdgeFlow: a technique for boundary detection and image segmentation," *IEEE Trans. Image Process.* **9**, 1375–1388 (2000).
14. D. Crisp and T. Tao, "Fast region merging algorithms for image segmentation," in *5th Asia Conf. on Computer Vision*, Melbourne, 22–25 January 2002, pp. 412–417, Asian Federation of Computer Vision Societies, Tokyo (2002).
15. P. C. Hansen and D. P. O″Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Comput. (USA)* **14**, 1487–1503 (1993).
16. X. He, N. H. C. Yung, K. P. Chow, F. Y. L. Chin, R. H. Y. Chung, K. Y. K. Wong, and K. S. H. Tsang, "Watershed segmentation with boundary curvature ratio based merging criterion," in *Proc. 9th Signal Image Processing*, pp. 7–12, Acta Press, Calgary, Canada (2007).
17. J. Freixenet, X. Munoz, D. Raba, J. Marti, and X. Cufi, "Yet another survey on image segmentation: Region and boundary information integration," *Proc. 7th European Conf. on Computer Vision*, *Lect. Notes Comput. Sci.* **2352**, 408–422 (2002).
18. K. Haris, "Hybrid image segmentation using watersheds and fast region merging," *IEEE Trans. Image Process.* **7**, 1684–1699 (1998).
19. T. Pavlidis and Y. T. Liow, "Integrating region growing and edge-detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **12**, 225–233 (1990).
20. S. E. Hernandez and K. E. Barner, "Joint region merging criteria for watershed-based image segmentation," *Proc. Int. Conf. on Image Processing (ICIP)*, Vol. **2**, pp. 108–111, IEEE, Piscataway, NJ (2000).
21. D. M. Wang, "A multiscale gradient algorithm for image segmentation using watersheds," *Pattern Recogn.* **30**, 2043–2052 (1997).
22. R. H. Y. Chung, N. H. C. Yung, and P. Y. S. Cheung, "An efficient parameterless quadrilateral-based image segmentation method," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**, 1446–1458 (2005).
23. X. Ren and J. Malik, "Learning a classification model for segmentation," *Proc. Int. Conf. on Image Processing (ICIP)*, Vol. **1**, pp. 10–17, IEEE, Piscataway, NJ (2003).
24. M. A. Ruzon and C. Tomasi, "Color edge detection with the compass operator," *Proc. Computer Vision and Pattern Recognition (CVPR)*, Vol. **2**, pp. 166, IEEE, Piscataway, NJ (1999).
25. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice Hall, Englewood Cliffs, NJ (2002).
26. J. Canny, "A computational approach to edge-detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **8**, 679–698 (1986).
27. X. He and N. H. C. Yung, "Performance improvement of edge detection based on edge likelihood index," in *Visual Communications and Image Processing (VCIP)*, *Proc. SPIE* **5960**, 59604W (2005).
28. G. Koepfler, C. Lopez, and J. M. Morel, "A multiscale algorithm for image segmentation by variational method," *SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal.* **31**, 282–299 (1994).
29. A. Brook, R. Kimmel, and N. A. Sochen, "Geometry motivated variational segmentation for color images," *Scale-Space Morphol. Computer Vision Proc.*, *Lect. Notes Comput. Sci.* **2106**, 362–370 (2001).
30. E. H. Sergio, E. B. Kenneth, and Y. Yu, "Region merging using homogeneity and edge integrity for watershed-based image segmentation," *Opt. Eng.* **44**, 017004 (2005).

Zhi-Gang Tan received his MEng degree from Zhejiang University. He is pursuing his PhD degree in the Department of Electrical and Electronic Engineering, University of Hong Kong. His research interests include pattern recognition, computer vision, especially theories of image segmentation and classification, boundary detection, and statistical models.

Nelson H. C. Yung received his BSc and PhD degrees from the University of Newcastle-Upon-Tyne, where he was a lecturer from 1985 to 1990. From 1990 to 1993, he worked as senior research scientist at the Department of Defense, Australia. He joined the University of Hong Kong in late 1993 as an associate professor. He has coauthored five books and book chapters, and has published more than 150 journal and conference papers in the areas of digital image processing, parallel algorithms, visual traffic surveillance, autonomous vehicle navigation, and learning algorithms.