Color Image Segmentation Based on Mean Shift and Normalized Cuts

Wenbing Tao, Hai Jin, *Senior Member, IEEE*, and
Yimin Zhang, *Senior Member, IEEE*

Abstract—In this correspondence, we develop a novel approach that provides effective and robust segmentation of color images. By incorporating the advantages of the mean shift (MS) segmentation and the normalized cut (Ncut) partitioning methods, the proposed method requires low computational complexity and is therefore very feasible for real-time image segmentation processing. It preprocesses an image by using the MS algorithm to form segmented regions that preserve the desirable discontinuity characteristics of the image. The segmented regions are then represented by using the graph structures, and the Ncut method is applied to perform globally optimized clustering. Because the number of the segmented regions is much smaller than that of the image pixels, the proposed method allows a low-dimensional image clustering with significant reduction of the complexity compared to conventional graph-partitioning methods that are directly applied to the image pixels. In addition, the image clustering using the segmented regions, instead of the image pixels, also reduces the sensitivity to noise and results in enhanced image segmentation performance. Furthermore, to avoid some inappropriate partitioning when considering every region as only one graph node, we develop an improved segmentation strategy using multiple child nodes for each region. The superiority of the proposed method is examined and demonstrated through a large number of experiments using color natural scene images.

Index Terms—Color image segmentation, graph partitioning, mean shift (MS), normalized cut (Ncut).

I. INTRODUCTION

Image segmentation is a process of dividing an image into different regions such that each region is nearly homogeneous, whereas the union of any two regions is not. It serves as a key in image analysis and pattern recognition and is a fundamental step toward low-level vision, which is significant for object recognition and tracking, image retrieval, face detection, and other computer-vision-related applications [1]. Color images carry much more information than gray-level ones [24]. In many pattern recognition and computer vision applications, the color information can be used to enhance the image analysis process and improve segmentation results compared to gray-scale-based approaches. As a result, great efforts have been made in recent years to investigate segmentation of color images due to demanding needs.

Existing image segmentation algorithms can be generally classified into three major categories, i.e., feature-space-based clustering, spatial segmentation, and graph-based approaches. Feature-space-based clustering approaches [12], [13] capture the global characteristics of the image through the selection and calculation of the image features, which are usually based on the color or texture. By using a specific distance measure that ignores the spatial information, the feature

samples are handled as vectors, and the objective is to group them into compact, but well-separated clusters [7].

Although the data clustering approaches are efficient in finding salient image features, they have some serious drawbacks as well. The spatial structure and the detailed edge information of an image are not preserved, and pixels from disconnected regions of the image may be grouped together if their feature spaces overlap. Given the importance of edge information, as well as the need to preserve the spatial relationship between the pixels on the image plane, there is a recent tendency to handle images in the spatial domain [11], [28]. The spatial segmentation method is also referred to as region-based when it is based on region entities. The watershed algorithm [19] is an extensively used technique for this purpose. However, it may undesirably produce a very large number of small but quasi-homogenous regions. Therefore, some merging algorithm should be applied to these regions [20], [28].

Graph-based approaches can be regarded as image perceptual grouping and organization methods based on the fusion of the feature and spatial information. In such approaches, visual group is based on several key factors such as similarity, proximity, and continuation [3], [5], [21], [25]. The common theme underlying these approaches is the formation of a weighted graph, where each vertex corresponds to n image pixel or a region, and the weight of each edge connecting two pixels or two regions represents the likelihood that they belong to the same segment. The weights are usually related to the color and texture features, as well as the spatial characteristic of the corresponding pixels or regions. A graph is partitioned into multiple components that minimize some cost function of the vertices in the components and/or the boundaries between those components. So far, several graph cut-based methods have been developed for image segmentations [8], [14], [22], [23], [27], [30], [31]. For example, Shi and Malik [23] proposed a general image segmentation approach based on normalized cut (Ncut) by solving an eigensystem, and Wang and Siskind [8] developed an image-partitioning approach by using a complicated graph reduction. Besides graph-based approaches, there are also some other types of image segmentation approaches that mix the feature and spatial information [4], [29].

This correspondence concerns a Ncut method in a large scale. It has been empirically shown that the Ncut method can robustly generate balanced clusters and is superior to other spectral graph-partitioning methods, such as average cut and average association [23]. The Ncut method has been applied in video summarization, scene detection [17], and cluster-based image retrieval [18]. However, image segmentation approaches based on Ncut, in general, require high computation complexity and, therefore, are not suitable for real-time processing [23]. An efficient solution to this problem is to apply the graph representation strategy on the regions that are derived by some region segmentation method. For example, Makrogiannis *et al.* [20] developed an image segmentation method that incorporates region-based segmentation and graph-partitioning approaches. This method first produces a set of oversegmented regions from an image by using the watershed algorithm, and a graph structure is then applied to represent the relationship between these regions.

Not surprisingly, the overall segmentation performance of the region-based graph-partitioning approaches is sensitive to the region segmentation results and the graph grouping strategy. The inherent oversegmentation effect of the watershed algorithm used in [20] and [28] produces a large number of small but quasi-homogenous regions, which may lead to a loss in the salient features of the overall image and, therefore, yield performance degradation in the consequent region grouping.

To overcome these problems, we propose in this correspondence a novel approach that provides effective and robust image segmentation

with low computational complexity by incorporating the mean shift (MS) and the Ncut methods. In the proposed method, we first perform image region segmentation by using the MS algorithm [4], and we then treat these regions as nodes in the image plane and apply a graph structure to represent them. The final step is to apply the Ncut method to partition these regions.

The MS algorithm is a robust feature-space analysis approach [4] which can be applied to discontinuity preserving smoothing and image segmentation problems. It can significantly reduce the number of basic image entities, and due to the good discontinuity preserving filtering characteristic, the salient features of the overall image are retained. The latter property is particularly important in the partitioning of natural images, in which only several distinct regions are used in representing different scenes such as sky, lake, sand beach, person, and animal, whereas other information within a region is often less important and can be neglected. However, it is difficult to partition a natural image into significative regions to represent distinct scenes, depending only on the MS segmentation algorithm. The main reason is that the MS algorithm is an unsupervised clustering-based segmentation method, where the number and the shape of the data cluster are unknown *a priori*. Moreover, the termination of the segmentation process is based on some region-merging strategy applied to the filtered image result, and the number of regions in the segmented image is mainly determined by the minimum number of pixels in a region, which is denoted as M (i.e., regions containing less than M pixels will be eliminated and merged into its neighboring region). In our proposed approach, therefore, an appropriate value of M is chosen to yield a good region representation in the sense that the number of segmented regions is larger than the number of typical scenes, but is much smaller than the number of pixels in the image, and the boundary information of the typical scenes is retained by the boundaries of the regions.

The Ncut method [23], on the other hand, can be considered as a classification method. In most image segmentation applications, the Ncut method is applied directly to the image pixels, which are typically of very large size and thus require huge computational complexity. For example, to use the Ncut method in [26], a gray image has to be decimated into a size of 160×160 pixels or smaller. In summary, it is difficult to get real-time segmentation using the Ncut method. In the proposed method, the Ncut method is applied to the segmented regions instead of the raw image pixels. As such, it eliminates the major problem of the Ncut method that requires prohibitively high complexity. By applying the Ncut method to the preprocessed regions rather than the raw image pixels, the proposed method achieves a significant reduction of the computational cost and, therefore, renders real-time image segmentation much more practically implemental. On the other hand, due to some approximation in the implementation of the Ncut method, the segmentation processing of a graph exploiting the lower dimensional region-based weight matrix also provides more precise and robust partitioning performance compared to that based on the pixel-based weight matrix.

This correspondence is organized as follows. In Section II, we introduce the background of the proposed method. In Section III, the proposed algorithm is described for effective color image partitioning. In Section IV, we demonstrate the superiority of the proposed method by comparing the performance of the proposed approach to existing methods using a variety of color natural scene images. Finally, Section V concludes this correspondence.

II. MS AND GRAPH PARTITIONING

A. Image Region Segmentation Based on MS

The computational module based on the MS procedure is an extremely versatile tool for feature-space analysis. In [4], two applica-

tions of the feature-space analysis technique are discussed based on the MS procedure: discontinuity preserving filtering and the segmentation of gray-level or color images. In this section, we present a brief review of the image segmentation method based on the MS procedure [4], [9], [10].

We consider radially symmetric kernels satisfying $K(x) = c_{k,d}k(\|x\|^2)$, where constant $c_{k,d} > 0$ is chosen such that $\int_0^\infty K(x)dx = \int_0^\infty c_{k,d}k(\|x\|^2)dx = 1$ [note that $k(x)$ is defined only for $x \geq 0$]. $k(x)$ is a monotonically decreasing function and is referred to as the profile of the kernel. Given the function $g(x) = -k'(x)$ for profile, the kernel $G(x)$ is defined as $G(x) = c_{g,d}g(\|x\|^2)$. For n data points x_i, $i = 1, \ldots, n$, in the d-dimensional space R^d, the MS is defined as

$$m_{h,G}(x) = \frac{\sum_{i=1}^{n} x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x \qquad (1)$$

where x is the center of the kernel (window), and h is a bandwidth parameter. Therefore, the MS is the difference between the weighted mean, using kernel G as the weights and x as the center of the kernel (window). The MS method is guaranteed to converge to a nearby point where the estimate has zero gradient [4]. Regions of low-density values are of no interest for the feature-space analysis, and in such regions, the MS steps are large. On the other hand, near local maxima, the steps are small, and the analysis is more refined. The MS procedure, thus, is an adaptive gradient ascent method [6]. The center position of kernel G can be updated iteratively by

$$y_{j+1} = \frac{\sum_{i=1}^{n} x_i g\left(\left\|\frac{y_j-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{y_j-x_i}{h}\right\|^2\right)}, \qquad j = 1, 2, \ldots \qquad (2)$$

where y_1 is the center of the initial position of the kernel.

Based on the above analysis, the MS image filtering algorithm can be obtained. First, an image is represented as a 2-D lattice of p-dimensional vectors (pixels), where $p = 1$ for gray-level images, $p = 3$ for color images, and $p > 3$ for multispectral images. The space of the lattice is known as the spatial domain, while the graph level and the color of spectral information are represented in the range domain. For both domains, the Euclidean metric is assumed. Let x_i and z_i, $i = 1, \ldots, n$, respectively, be the d-dimensional ($d = p + 2$) input and the filtered image pixels in the joint spatial-range domain.

The segmentation is actually a merging process performed on a region that is produced by the MS filtering. The use of the MS segmentation algorithm requires the selection of the bandwidth parameter $\mathbf{h} = (h_r, h_s)$, which, by controlling the size of the kernel, determines the resolution of the mode detection.

B. Spectral Graph Partitioning

Among many graph theoretic algorithms, spectral graph-partitioning methods have been successfully applied to many areas in computer vision, including motion analysis [16], image segmentation [8], [23], [27], [31], image retrieval [18], video summarization [17], and object recognition [15]. In this correspondence, we use one of these techniques, namely, the Ncut method [23], for region clustering. Next, we briefly review the Ncut method based on the studies in [14], [23], and [27].

Roughly speaking, a graph-partitioning method attempts to organize nodes into groups such that the intragroup similarity is high and the
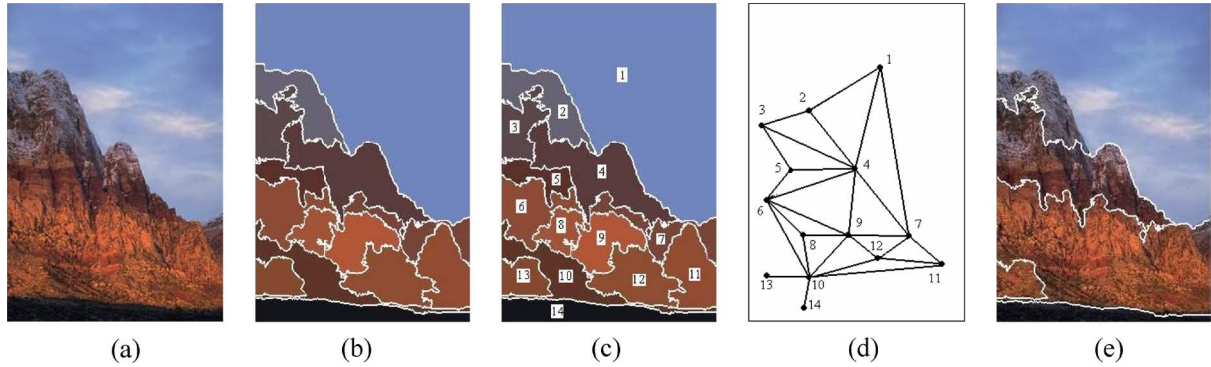
Fig. 1. (a) Original image. (b) Result image after using the MS segmentation algorithm. (c) Labeled regions. (d) RAGs produced by the node space relation, with a region corresponding to a node. (e) Final region-partitioning result using the Ncut method on the RAG in (d).

intergroup similarity is low. Given a graph $\mathbf{G} = (\mathbf{V}, \mathbf{E}, \mathbf{W})$, where $\mathbf{V}$ is the set of nodes, and $\mathbf{E}$ is the set of edges connecting the nodes. A pair of nodes u and ν is connected by an edge and is weighted by $w(u, \nu) = w(\nu, u) \geq 0$ to measure the dissimilarity between them. $\mathbf{W}$ is an edge affinity matrix with $w(u, \nu)$ as its (u, ν)th element. The graph can be partitioned into two disjoint sets $\mathbf{A}$ and $\mathbf{B} = \mathbf{V} - \mathbf{A}$ by removing the edges connecting the two parts. The degree of dissimilarity between the two sets can be computed as a total weight of the removed edges. This closely relates to a mathematical formulation of a cut [22]

$$\text{cut}(\mathbf{A}, \mathbf{B}) = \sum_{u \in \mathbf{A}, \nu \in \mathbf{B}} w(u, \nu). \qquad (3)$$

This problem of finding the minimum cut has been well studied [22], [23], [27]. However, the minimum cut criterion favors grouping small sets of isolated nodes in the graph because the cut defined in (3) does not contain any intragroup information [23]. In other words, the minimum cut usually yields overclustered results when it is recursively applied. This motivates several modified graph partition criteria, including the Ncut [23]

$$\text{Ncut}(\mathbf{A}, \mathbf{B}) = \frac{\text{cut}(\mathbf{A}, \mathbf{B})}{\text{assoc}(\mathbf{A}, \mathbf{V})} + \frac{\text{cut}(\mathbf{A}, \mathbf{B})}{\text{assoc}(\mathbf{B}, \mathbf{V})} \qquad (4)$$

where $\text{assoc}(\mathbf{A}, \mathbf{V})$ denotes the total connection from nodes in $\mathbf{A}$ to all nodes in the graph, and $\text{asso}(\mathbf{B}, \mathbf{V})$ is similarly defined. Unlike the cut criterion that has a bias in favor of cutting small sets of nodes, the Ncut criterion is unbiased.

III. PROPOSED APPROACH

A. Description of the Algorithm Scheme

We now describe our proposed algorithm. From a data-flow point of view, the outline of the proposed algorithm can be characterized as the following. First, an image is segmented into multiple separated regions using the MS algorithm. Second, the graph representation of these regions is constructed, and the dissimilarity measure between the regions is defined. Finally, a graph-partitioning algorithm based on the Ncut is employed to form the final segmentation map.

The regions produced by the MS segmentation can be represented by a planar weighted region adjacency graph (RAG) $\mathbf{G} = (\mathbf{V}, \mathbf{E}, \mathbf{W})$ that incorporates the topological information of the image structure and region connectivity. The majority of region-merging algorithms define the region dissimilarity metric as the distance between two adjacent regions in an appropriate feature space. This dissimilarity metric plays a decisive role in determining the overall performance of

the image segmentation process. To define the measure of dissimilarity between neighboring regions, we first define an appropriate feature space. Features like color, texture, statistical characteristics, and 2-D shape are useful for segmentation purposes and can be extracted from an image region. We adopt the color feature in this paper because it is usually the most dominant and distinguishing visual feature and adequate for a number of segmentation tasks. The average color components are computed over a region's pixels and are described by a three-element color vector. When an image is segmented based on the MS method into n regions R_i, $i = 1, \ldots, n$, the mean vector $\bar{X}_{R_i} = \{\bar{x}_{1i}, \bar{x}_{2i}, \bar{x}_{3i}\}$ is computed for each region, where $\bar{x}_{1i}$, $\bar{x}_{2i}$, and $\bar{x}_{3i}$ are the mean pixel intensities of the ith region in the three different color spaces, respectively.

Proper selection of the color spaces is important to the development of a good region-merging algorithm. To obtain meaningful segmentation results, the perceived color difference should be associated with the Euclidean distance in the color space. The spaces $\mathbf{L}^* \mathbf{u}^* \mathbf{v}^*$ and $\mathbf{L}^* \mathbf{a}^* \mathbf{b}^*$ have been particularly designed to closely approximate the perceptually uniform color spaces. In both cases, $\mathbf{L}^*$, which is the lightness (relative brightness) coordinate, is defined in the same way. The two spaces differ only in the chromaticity coordinates, and in practice, there is no clear advantage of using one over the other. In this paper, we employed $\mathbf{L}^* \mathbf{u}^* \mathbf{v}^*$ motivated by its linear mapping property [2].

By defining the color space, we can compute the weight matrix $\mathbf{W}$ of all regions. The weight $w(u, \nu)$ between regions u and ν is defined as

$$w(u, \nu) = \begin{cases} e^{-\left[\frac{\|\mathbf{F}(u) - \mathbf{F}(\nu)\|_2^2}{d_I} \right]}, & \text{if } u \text{ and } \nu \text{ are adjacent} \\ 0, & \text{otherwise} \end{cases} \qquad (5)$$

where $\mathbf{F}(u) = \{\mathbf{L}(u), \mathbf{u}(u), \mathbf{v}(u)\}$ is the color vector of region u, and $\| \cdot \|_2$ denotes the vector norm operator. In addition, d_I is a positive scaling factor that determines the sensitivity of $w(u, \nu)$ to the color difference between nodes u and ν.

Under a graph representation, region grouping can be naturally formulated as a graph-partitioning problem. In the proposed method, the Ncut algorithm is used to solve such a problem. The major difference between the proposed method and the conventional Ncut algorithm is that the construction of the weight matrix is based not on the pixels of the original image but rather on the segmentation result of the MS algorithm. The advantages of using regions instead of pixels in constructing the weight matrix are twofold: 1) It offers a considerable reduction of computational complexity, since the number of basic image entities is by far smaller than that of the pixels. Thus, the size of the weight matrix and, subsequently, the complexity of the graph

TABLE I
WEIGHT MATRIX **W** OF ALL REGION NODES

Regions	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	0.30	0	0.02	0	0	0.02	0	0	0	0	0	0	0
2	0.30	1	0.63	0.43	0	0	0	0	0	0	0	0	0	0
3	0	0.63	1	0.94	0.77	0	0	0	0	0	0	0	0	0
4	0.02	0.43	0.94	1	0.93	0.59	0.84	0	0.22	0	0	0	0	0
5	0	0	0.77	0.93	1	0.54	0	0	0	0	0	0	0	0
6	0	0	0	0.59	0.54	1	0	0.91	0.77	0.58	0	0	0	0
7	0.02	0	0	0.84	0	0	1	0	0.51	0	0.87	0.89	0	0
8	0	0	0	0	0	0.91	0	1	0.95	0.34	0	0	0	0
9	0	0	0	0.22	0	0.77	0.51	0.95	1	0.21	0	0.78	0	0
10	0	0	0	0	0	0.58	0	0.34	0.21	1	0.54	0.56	0.70	0.29
11	0	0	0	0	0	0	0.87	0	0	0.54	1	1	0	0
12	0	0	0	0	0	0	0.89	0	0.78	0.56	1	1	0	0
13	0	0	0	0	0	0	0	0	0	0.70	0	0	1	0
14	0	0	0	0	0	0	0	0	0	0.29	0	0	0	1

TABLE II
WEIGHT MATRIX **W** OF ALL THE CHILD NODES OF REGIONS 1–4

|											
1	1	1	0.30	0.30	0.30	0	0	0	0.02	0.02	0.02
1	1	1	0.30	0.30	0.30	0	0	0	0.02	0.02	0.02
1	1	1	0.30	0.30	0.30	0	0	0	0.02	0.02	0.02
0.30	0.30	0.30	1	1	1	0.63	0.63	0.63	0.43	0.43	0.43
0.30	0.30	0.30	1	1	1	0.63	0.63	0.63	0.43	0.43	0.43
0.30	0.30	0.30	1	1	1	0.63	0.63	0.63	0.43	0.43	0.43
0	0	0	0.63	0.63	0.63	1	1	1	0.94	0.94	0.94
0	0	0	0.63	0.63	0.63	1	1	1	0.94	0.94	0.94
0	0	0	0.63	0.63	0.63	1	1	1	0.94	0.94	0.94
0.02	0.02	0.02	0.43	0.43	0.43	0.94	0.94	0.94	1	1	1
0.02	0.02	0.02	0.43	0.43	0.43	0.94	0.94	0.94	1	1	1
0.02	0.02	0.02	0.43	0.43	0.43	0.94	0.94	0.94	1	1	1

structure employed for image representation are significantly reduced. As a result, the bottleneck problem of computation and storage requirements in applying the Ncut algorithm is solved. 2) It achieves improved segmentation performance. Although the implementation strategy of using the Ncut method to partition a graph based on region nodes is the same as that based on pixel nodes, the feature difference between the region nodes formed by the MS algorithm is larger and more robust than that between the pixel nodes. Therefore, the partitioning of the graph based on regions using the Ncut method is easier and more accurate than that based on pixels. Moreover, the MS algorithm not only removes noise, which limits the accuracy of graph partitioning in the Ncut method, but also adaptively reduces the amount of smoothing near abrupt changes in the edges and, as such, retains the salient features of the overall image.

B. Illustration of the Implementation Procedure

In order to illustrate the implementation process of the proposed algorithm, we use the example of a mountain image under sky, as depicted in Fig. 1. The image size is 256×384. Fig. 1(a) shows the original image, and Fig. 1(b) depicts the resultant image after applying the MS segmentation algorithm, with the white contours depicting the boundaries between the regions. We set $\mathbf{h} = (h_r, h_s) = (6, 8)$ and $M = 2000$. As a result, the MS segmentation algorithm produces 14 regions. Fig. 1(c) shows the labeled region result, and Fig. 1(d) is the RAG derived from the labeled regions, where only the neighboring regions have weighted edges. Then, the weight matrix **W** of all region nodes is computed, and the results are summarized in Table I. Fig. 1(e) depicts the region-partitioning result using the Ncut algorithm, with partitioning class $k = 4$. The resulting four eigenvalues, the corresponding eigenvectors, and the discretized eigenvectors are used to partition the region nodes. Based on the discrete eigenvectors, the following four clusters are formed: Regions 1 and 2 are merged to
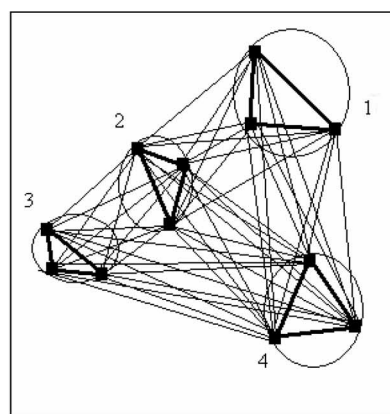
(a)
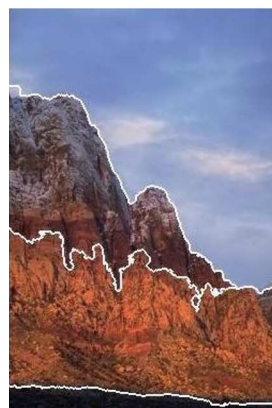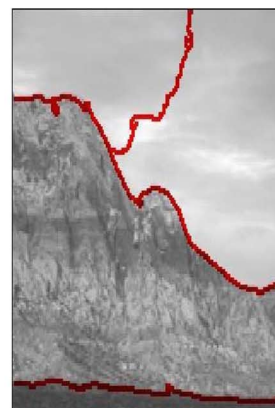
(b)   (c)

Fig. 2. (a) RAGs produced by the node space relations, with a region corresponding to three nodes. (b) Final region-partitioning result using the Ncut method on the RAG in (a). (c) Partitioning result using the Ncut method implemented by Cour *et al.* [26].

form the first cluster, regions 3–5 the second cluster, regions 6–12 the third cluster, and regions 13 and 14 the fourth cluster.

C. Segmentation Improvement Using Multiple Child Nodes

Although the graph-partitioning process merges the regions generated by the MS algorithm into four clusters, the final region segmentation result in Fig. 1(e) is not perfect because the sky and the top region of the mountain near the sky are merged into the same cluster. The grouping of the other regions is not satisfactory either. It is well
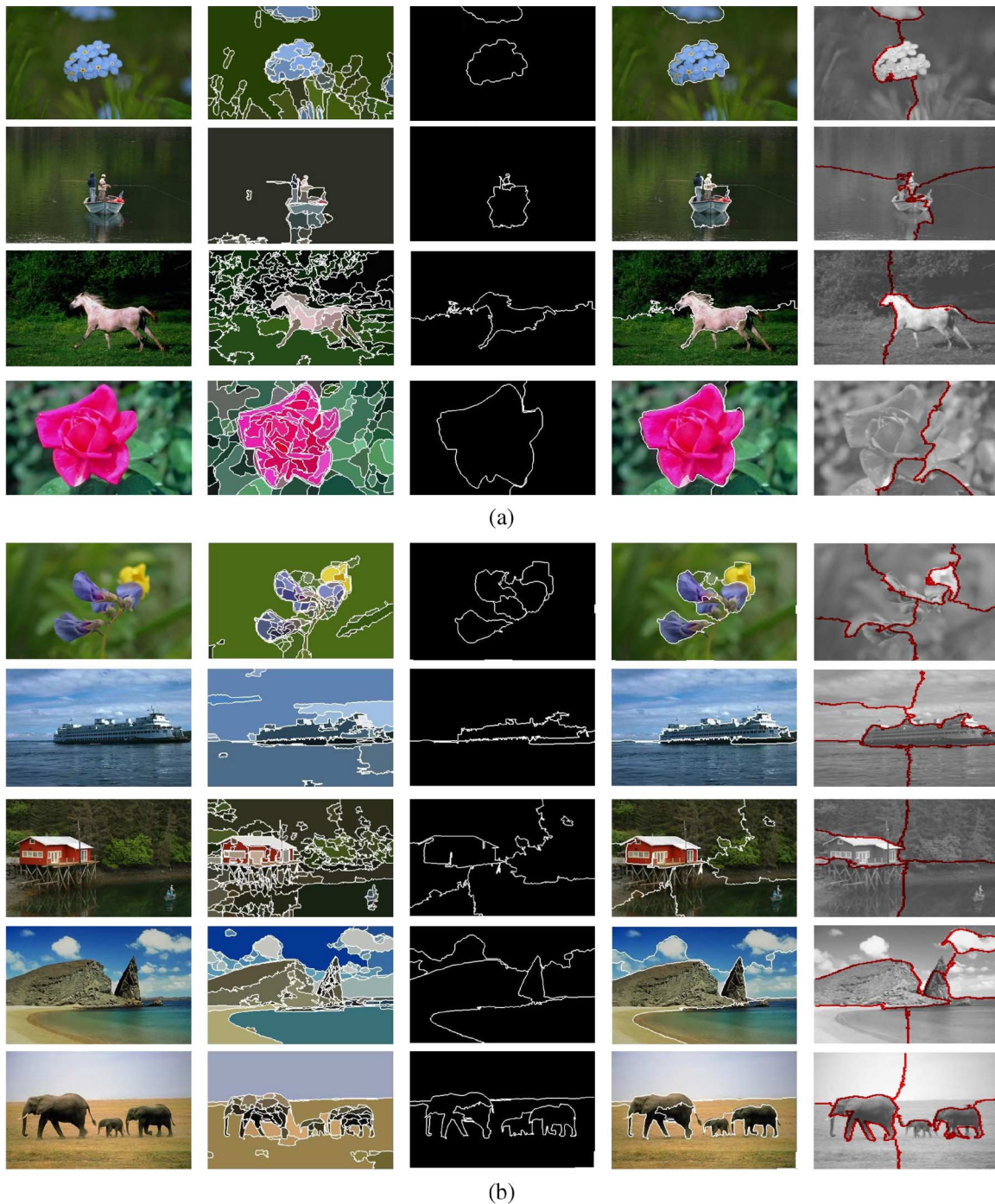
Fig. 3. Test results of color images with partitioning class $k = 3$ in (a) and $k = 6$ in (b). (First column) The original test images. (Second column) The region segmentation results by the MS algorithm. The white lines show the contour boundaries of these regions, and the number of regions of the result image in each row of (a) and (b) (from top to bottom) is 90, 28, 104, 106, 90, 38, 159, 97, and 66, respectively. (Third column) The contour images of the final region-merging results using the proposed method. (Fourth column) The original images overlapped with the contour of the final region-partitioning results. (Fifth column) The partitioning results by directly applying the Ncut method to the image pixels [26]. The image size is 240×160.

known that the Ncut algorithms try to find a "balanced partition" [23] of a weighted graph, and the cut does not have a bias in favor of cutting small sets of isolated nodes in the graph. Therefore, it is not surprising that the sky cannot be segmented out because it is only considered as a node of the weighted graph.

An effective solution to this problem is to consider every region generated by the MS algorithm as multiple nodes rather than a single one. The child nodes of a region have the same feature value and are all adjacent with all the child nodes of the neighboring region. That is

to say, there are edges between all the child nodes of two adjacent regions. The weights between the child nodes within a region are all one, whereas the weights between the child nodes between two adjacent regions are all the same and equal to the weight between the two regions. This yields a new weight matrix $W_C = W \otimes 1_C$, where $\otimes$ denotes the Kronecker product operator, and $\mathbf{1}_C$ is the $C \times C$ matrix with all unit entries. In this experiment, we set the number of child nodes in each region to $C = 3$, and the top-left 12×12 entries of $\mathbf{W}_C$, which correspond to regions 1–4 as shown in bold fonts in
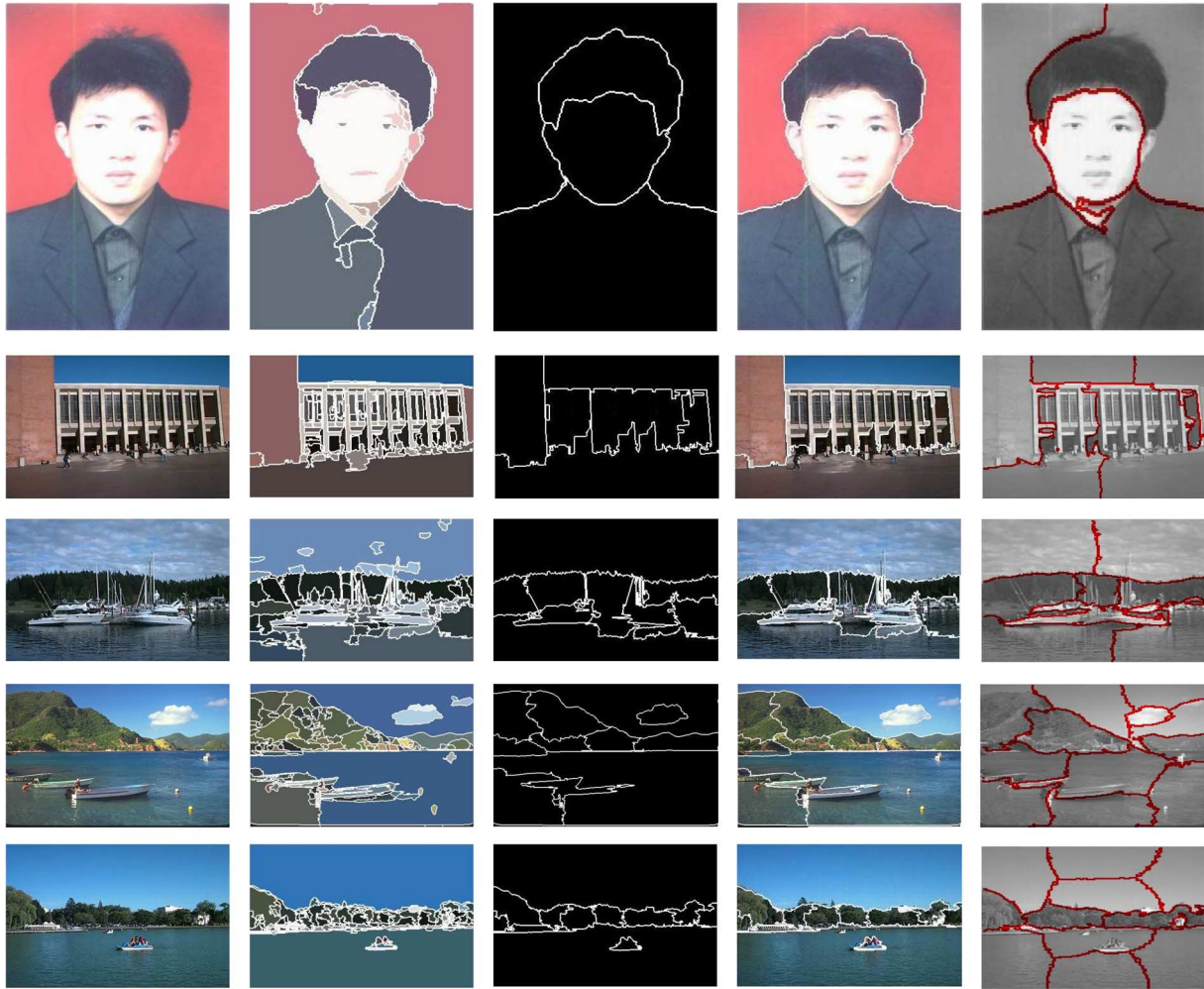
Fig. 4. Test results of color images with the variable partitioning class. The partitioning class k of each row (from top to bottom) is 4, 6, 8, 10, and 12, respectively. The number of regions of the result image in the second column is (from top to bottom) 49, 117, 92, 128, and 98, respectively. The size of the image is 160×240 for the first row and 240×160 for other rows.

TABLE III
COMPARED COMPUTATIONAL COST BETWEEN THE PROPOSED METHOD AND THE Ncut METHOD

Image No. (from top to bottom)		Region num. of MS	Time of MS (s)	Time of *Ncut* Merging (s)	Final region num.	Time of *Ncut* method [26] (s)
As in Fig. 3(a)	1	90	2.30	0.101		36.406
	2	28	2.31	0.027		30.344
	3	104	2.54	0.125	3	41.969
	4	106	1.59	0.139		39.922
As in Fig. 3(b)	1	90	2.70	0.132		31.968
	2	38	2.14	0.047		34.500
	3	159	2.31	0.232	6	34.657
	4	97	1.90	0.141		41.078
	5	66	1.93	0.093		31.906
As in Fig. 5	1	49	1.96	0.063	4	32.813
	2	117	1.64	0.203	6	36.000
	3	92	2.60	0.141	8	40.781
	4	128	2.53	0.266	10	42.359
	5	98	2.18	0.203	12	44.844

Table I, are depicted in Table II. The adjacent graph of all the child nodes of regions 1–4 is shown in Fig. 2(a).

Fig. 2(b) illustrates the final region-partitioning result using the Ncut algorithm based on the new weight matrix $\mathbf{W}_C$, where the partitioning class remains $k = 4$. In the modified scheme, region 1 alone forms a cluster, regions 2–5 are merged to form the second cluster, regions 6–13 as the third cluster, and region 14 forms the fourth cluster. Note that the sky and the mountain are now separated into different clusters. Evidently, the final region segmentation result shown in Fig. 2(b) is better than that shown in Fig. 1(e). For comparison, Fig. 2(c) also

shows the segmentation results of the Ncut method implemented by Cour *et al.* [26] where the sky is separated into two parts.

IV. EXPERIMENTAL RESULTS

We have applied the proposed algorithm for the segmentation of a set of color images with natural scenes. In this section, we present the experimental results, indicating the different stages of the method. The contrastive experiment results using the Ncut method implemented by Cour *et al.* [26] are also presented for comparison. The sizes of all the test images are either 240×160 or 160×240. The parameters of the MS segmentation algorithm are set to $\mathbf{h} = (h_r, h_s) = (6, 8)$ and $M = 50$. With this parameter setting, the number of the regions produced by the MS algorithm is less than 200 regions (i.e., the RAG has less than 200 nodes) for all the test images we used. Therefore, the dimension of the weight matrix is dramatically reduced from the number of image pixels to the number of regions, and subsequently, the complexity of the graph structure employed for image representation is also reduced. To retain the advantage of the balanced partition of the Ncut method, we use $C = 3$ to construct the weight matrix $\mathbf{W}_C$.

The test examples include three image sets. The results of the first set of examples are shown in Fig. 3(a), where the partitioning class k is three. The results of the second set of examples are shown in Fig. 3(b), where $k = 6$. The results of the third set of examples are shown in Fig. 4 with varying values of k. In each figure, the five columns respectively show, from left to right, the original test image, the resultant image after applying the MS segmentation algorithm, the contour image of the final region partitioning, the original images overlapped with the final partitioning contour, and the partitioning result using the Ncut method implemented by Cour *et al.* [26].

As discussed in Section III-C, considering a region produced by the MS algorithm as several identical child nodes can prevent the pixels in this region from being divided into several parts when the Ncut method is applied. For example, in the second row of Fig. 3(a), the image includes the fisher in the foreground and the lake in the background. Therefore, an effective segmentation method should distinguish the fisher from the lake background. However, when the Ncut method is directly applied to the image pixels, the image is partitioned into three regions where each region includes one part of the lake. If we first process the image using the MS segmentation algorithm, then the most pixels of the lake background are formed into a region. The proposed method treats this region as several nodes with the identical characteristic and, as such, avoids partitioning the lake into several different parts. The experimental results of the second and fifth rows in Fig. 3(b) and the third and fifth rows in Fig. 4 also support the same claim.

For all the experimental results shown in Figs. 3 and 4, the proposed method effectively partitions the natural scenes into several meaningful regions and provides an improved performance compared to the Ncut method, which does not always yield meaningful regions. Fig. 3(a) shows images with relatively simple scene, and thus, a small partitioning class of $k = 3$ is chosen. The test sets in Figs. 3(b) and 4 include more complicated color natural scenes, and thus, a larger value of k is used.

Finally, we consider the computational cost of the proposed method and compare it with that of the Ncut method. A PC, which is equipped with a 2.0-GHz Pentium CPU and 512-MB memory, is used. The main computational cost of the Ncut method is to compute the eigenvectors of the source image. In the experiments used in this paper, the Ncut method [26] requires 30–50 s to compute the eigenvectors of the gray image of 240×160 pixels. The computation of the proposed method consists of two sections. The first section is to segment the original image using the MS segmentation algorithm, which takes approximately 2 s to process an image. The second section is to partition the region

nodes produced by the MS segmentation algorithm using the Ncut method, which depends on the number of the initial region nodes and the requested final regions. In the experiments performed in this paper, the number of the regions is less than 200, and for $C = 3$, the number of the child nodes is less than 600. The computation of partitioning the regions using the Ncut method takes less than 300 ms. Table III compares the computational cost between the proposed method and the Ncut method for the images depicted in Figs. 3 and 4. The significant reduction of the computational cost by using the proposed method is evident.

Moreover, we notice that the MS algorithm is more feasible for parallel operations than the Ncut method because the complicated eigendecomposition problem in the Ncut method is difficult to be parallelized. In the proposed method, the main computational cost is the region segmentation, whereas the computational cost of partitioning the region nodes using the Ncut method is negligibly small. Thus, the proposed method is more feasible to real-time image processing, such as content-based image retrieval, particularly when parallel processing is used.

V. CONCLUSION

In this correspondence, we have developed a new algorithm for the segmentation of color images. The proposed algorithm takes the advantages of the MS segmentation method and the Ncut grouping method, whereas their drawbacks are avoided. The use of the MS method permits the formation of segments that preserve discontinuity characteristic of an image. On the other hand, the application of the region adjacent graph and Ncut methods to the resulting segments, rather than directly to the image pixels, yields superior image segmentation performance. The proposed method requires significantly lower computational complexity and, therefore, is feasible to real-time image processing.

REFERENCES

[1] N. Pal and S. Pal, "A review on image segmentation techniques," *Pattern Recognit.*, vol. 26, no. 9, pp. 1277–1294, Sep. 1993.
[2] G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. New York: Wiley, 1982.
[3] G. Kanizsa, *Grammatica del Vedere*. Bologna, Italy, 1980.
[4] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
[5] B. Caselles, B. Coll, and J.-M. Miorel, "A Kanisza programme," *Progr. Nonlinear Differential Equations Appl.*, vol. 25, pp. 35–55, 1996.
[6] R. van den Boomgaard and J. van de Weijer, "On the equivalence of local-mode finding, robust estimation and mean-shift analysis as used in early vision tasks," in *Proc. Int Conf. Pattern Recog.*, 2002, pp. 30 927–30 930.
[7] R. O. Duda, P. E. Hart, and D. G. Sork, *Pattern Classification*. New York: Wiley-Interscience, 2000.
[8] S. Wang and J. M. Siskind, "Image segmentation with ratio cut," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 6, pp. 675–690, Jun. 2003.
[9] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, Aug. 1995.
[10] D. Comaniciu, "An algorithm for data-driven bandwidth selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 281–288, Feb. 2003.
[11] V. Grau, A. U. J. Mewes, M. Alcaniz, R. Kikinis, and S. K. Warfield, "Improved watershed transform for medical image segmentation using prior information," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 447–458, Apr. 2004.
[12] D. W. Jacobs, D. Weinshall, and Y. Gdalyahu, "Classification with non-metric distances: Image retrieval and class representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 6, pp. 583–600, Jun. 2000.

[13] A. K. Jain and D. Zongker, "Representation and recognition of handwritten digits using deformable templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 12, pp. 1386–1391, Dec. 1997.

[14] S. X. Yu and J. Shi, "Multiclass spectral clustering," in *Proc. Int. Conf. Comput. Vis.*, 2003, pp. 313–319.

[15] S. Sarkar and P. Soundararajan, "Supervised learning of large perceptual organization: Graph spectral partitioning and learning automata," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 5, pp. 504–525, May 2000.

[16] J. Costeira and T. Kanade, "A multibody factorization method for motion analysis," in *Proc. Int. Conf. Comput. Vis.*, 1995, pp. 1071–1076.

[17] C.-W. Ngo, Y.-F. Ma, and H.-J. Zhang, "Video summarization and scene detection by graph modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 2, pp. 296–305, Feb. 2005.

[18] Y. Chen, J. Z. Wang, and R. Krovetz, "CLUE: Cluster-based retrieval of images by unsupervised learning," *IEEE Trans. Image Process.*, vol. 14, no. 8, pp. 1187–1201, Aug. 2005.

[19] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–597, Jun. 1991.

[20] S. Makrogiannis, G. Economou, and S. Fotopoulos, "A region dissimilarity relation that combines feature-space and spatial information for color image segmentation," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 35, no. 1, pp. 44–53, Feb. 2005.

[21] M. Wertheimer, "Laws of organization in perceptual forms (partial translation)," in *Sourcebook of Gestalt Psychology*, W. B. Ellis, Ed. Orlando, FL: Harcourt Brace Jovanovich, 1938, pp. 71–88.

[22] Z.-Y. 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.*, vol. 15, no. 11, pp. 1101–1113, Nov. 1993.

[23] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[24] H. D. Cheng, X. H. Jiang, Y. Sun, and J. Wang, "Color image segmentation: Advances and prospects," *Pattern Recognit.*, vol. 34, no. 12, pp. 2259–2281, Dec. 2001.

[25] O. J. Morris, J. Lee, and A. G. Constantinides, "Graph theory for image analysis: An approach based on the shortest spanning tree," *Proc. Inst. Electr. Eng., F*, vol. 133, no. 2, pp. 146–152, 1986.

[26] T. Cour, S. Yu, and J. Shi, *Normalized cuts Matlab code* [Online]. Available: http://www.cis.upenn.edu/~jshi/software

[27] Y. Weiss, "Segmentation using eigenvectors: A unifying view," in *Proc. Int. Conf. Comput. Vis.*, 1999, pp. 957–982.

[28] K. Haris, S. N. Efstratiadis, N. Maglaveras, and A. K. Katsaggelos, "Hybrid image segmentation using watersheds and fast region merging," *IEEE Trans. Image Process.*, vol. 7, no. 12, pp. 1684–1699, Dec. 1998.

[29] S. C. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing, and Bayes/MDL for multi-band image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 9, pp. 884–900, Sep. 1996.

[30] B. Sumengen and B. S. Manjunath, "Graph partitioning active contours (GPAC) for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 509–521, Apr. 2006.

[31] I. H. Jermyn and H. Ishikawa, "Globally optimal regions and boundaries as minimum ratio cycles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1075–1088, Oct. 2001.