

Available online at www.sciencedirect.com



Pattern Recognition Letters 26 (2005) 1139-1156

Pattern Recognition Letters

www.elsevier.com/locate/patrec

# Seeded region growing: an extensive and comparative study

Jianping Fan <sup>a,\*</sup>, Guihua Zeng <sup>b</sup>, Mathurin Body <sup>c</sup>, Mohand-Said Hacid <sup>c</sup>

<sup>a</sup> Department of Computer Science, University of North Carolina, 9201 University City Boulevard, Charlotte, NC 28223, USA <sup>b</sup> Department of Electronic Engineering, Shanghai Jiaotong University, Shanghai 200030, China <sup>c</sup> LISI-UFR d'Informatique, Universite Claude Bernard Lyon 1, France

> Received 26 November 2003 Available online 7 December 2004

#### Abstract

Seeded region growing (SRG) algorithm is very attractive for semantic image segmentation by involving high-level knowledge of image components in the seed selection procedure. However, the SRG algorithm also suffers from the problems of pixel sorting orders for labeling and automatic seed selection. An obvious way to improve the SRG algorithm is to provide more effective pixel labeling technique and automate the process of seed selection. To provide such a framework, we design an automatic SRG algorithm, along with a boundary-oriented parallel pixel labeling technique and an automatic seed selection method. Moreover, a seed tracking algorithm is proposed for automatic moving object extraction. The region seeds, which are located inside the temporal change mask, are selected for generating the regions of moving objects. Experimental evaluation shows good performances of our technique on a relatively large variety of images without the need of adjusting parameters.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Seeded region growing; Automatic image segmentation; Seed tracking

## 1. Introduction

Automatic image segmentation is an essential process for most subsequent tasks, such as image description, recognition, retrieval and object-based image compression (Majunath et al., 2000; Kunt et al., 1987). Automatic image segmentation has

\* Corresponding author.

also become a key point of MPEG-4 and MPEG-7 standards for realizing the object-based image coding and content-based image description and retrieval. The general image segmentation problem involves the partitioning of a given image into a number of homogeneous regions according to a given critical. Thus, image segmentation can be considered as a pixel labeling process in the sense that all pixels that belong to the same homogeneous region are assigned the same label (Haris et al., 1998). The existing automatic image segmentation

E-mail address: jfan@uncc.edu (J. Fan).

<sup>0167-8655/\$ -</sup> see front matter @ 2004 Elsevier B.V. All rights reserved. doi:10.1016/j.patrec.2004.10.010

techniques can be classified into five approaches, namely, *thresholding techniques* (Lim and Lee, 1990; Sahoo et al., 1988; Pal and Pal, 1993), *boundary-based methods* (Kass et al., 1987; Palmer et al., 1996), *region-based methods* (Haralick and Shapiro, 1985; Chang and Li, 1994; Hijjatoleslami and Kittler, 1998; Adams and Bischof, 1994), *hybrid techniques* (Pavlidis and Liow, 1990; Haddon and Boyce, 1990; Chu and Aggarwal, 1993), and *clustering-based techniques* (Pappas, 1992; Shen et al., 1998).

Seeded region growing (SRG), that is introduced by (Adams and Bischof, 1994), is robust, rapid and free of tuning parameters. These characteristics allow implementation of a very good algorithm which could be applied to large variety of images. SRG is also very attractive for semantic image segmentation by involving the high-level knowledge of image components in the seed selection procedure. However, the SRG algorithm also suffers from the problems of automatic seed generation and pixel sorting orders for labeling (Mehnert and Jackway, 1997; Fan et al., 2001a). There are several potential approaches to improving the SRG algorithm:

• Scan order optimization: The original SRG algorithm uses sequential sorted list as data structure (Adams and Bischof, 1994). All pixels are put into the sequential sorted list according to their delta value. The authors in (Mehnert and Jackway, 1997) have confirmed that a different order of processing pixels leads to different final segmentation results. They also noticed two types of order dependencies. The first type is called inherent order dependencies, while the second is called implementation order dependencies. However, the unlabeled pixels may not be adjacent to all these selected seeds especially at the beginning of SRG procedure, thus the connection characteristics among the adjacent pixels should be used in the pixel labeling procedure. The objective of image segmentation is to label the adjacent (connected each other on pixel level) similar pixels with the same symbol. Since the region boundaries are used for defining the boundaries of different image components, we propose a boundary-oriented technique to accelerate the seeded pixel labeling procedure. Our boundary-oriented pixel labeling technique can also support a parallel SRG procedure.

- Automatic seed selection: The authors in (Fan et al., 2001a) have developed an automatic edge-oriented seed generation technique to automate SRG algorithm. The color edge detection technique is first performed to obtain the simplified geometric structures of a color image. The centroids of the neighboring labeled color edges are then taken as the initial seeds for region growing. However, the edge-oriented SRG algorithm may induce oversegmentation problem because the color edges may be overdetected for the texture images and thus result in redundant seeds. There are two reasonable approaches to improving this edge-oriented SRG algorithm: one is to perform a post-procedure of similarity-based region merging after the SRG procedure, and the other is to perform an image filtering procedure before the color edge detection procedure. In this paper, we will propose an automatic edge-oriented seed generation technique via image filtering.
- Temporal seed tracking: SRG is very attractive for content-based image database applications by involving the high-level knowledge of image objects in the segmentation procedure. However, automatic semantic image segmentation is an ill-defined problem because semantic objects do not usually correspond to homogeneous regions in color or texture (Deng and Manjunath, 2001). Automatic moving object extraction via seed tracking may be one reasonable solution of this problem (Grinias and Tziritas, 2001). In this paper, we propose an interesting seed tracking technique for automatic moving object extraction.

This paper is organized as follows. A brief review of SRG technique is given in Section 2. In Section 3, we propose three automatic SRG techniques (their major steps are shown in Fig. 1). A comparative study of those three automatic SRG techniques is also given. Section 4 describes a seed tracking technique for automatic moving object extraction. Section 5 introduces our techniques



Fig. 1. The major steps for the three automatic SRG techniques.

for semantic-sensitive salient object detection. We conclude in Section 6.

#### 2. Seeded region growing: brief review

Seeded region growing approach to image segmentation is to segment an image into regions with respect to a set of q seeds (Adams and Bischof, 1994). Given the set of seeds,  $S_1, S_2, \ldots, S_q$ , each step of SRG involves one additional pixel to one of the seed sets. Moreover, these initial seeds are further replaced by the centroids of these generated homogeneous regions,  $R_1, R_2, \ldots, R_q$ , by involving the additional pixels step by step. The pixels in the same region are labeled by the same symbol and the pixels in variant regions are labeled by different symbols. All these labeled pixels are called the *allocated pixels*, and the others are called the *unallocated pixels*. Let H be the set of all unallocated pixels which are adjacent to at least one of the labeled regions.

$$H = \left\{ (x, y) \notin \bigcup_{i=1}^{q} R_i | N(x, y) \cap \bigcup_{i=1}^{q} R_i \neq \emptyset \right\}$$
(1)

where N(x, y) is the second-order neighborhood of the pixel (x, y) as shown in Fig. 2.

For the unlabeled pixel  $(x, y) \in H$ , N(x, y) meets just one of the labeled image region  $R_i$  and define  $\varphi(x, y) \in \{1, 2, ..., q\}$  to be that index such that  $N(x, y) \cap R_{\varphi(x, y)} \neq \emptyset$ .  $\delta(x, y, R_i)$  is defined as the difference between the testing pixel at (x, y) and its adjacent labeled region  $R_i$ .  $\delta(x, y, R_i)$  is calculated as

$$\delta(x, y, R_i) = \left| g(x, y) - g(X_i^c, Y_i^c) \right| \tag{2}$$

where g(x, y) indicates the values of the three color components of the testing pixel (x, y),  $g(X_i^c, Y_i^c)$ represents the average values of three color

(x-1, y-1)	(x, y-1)	(x+1, y-1)
(x-1, y)	(x, y)	(x+1, y)
(x-1, y+1)	(x, y+1)	(x+1, y+1)

Fig. 2. The second-order neighborhood N(x, y) of current testing pixel at (x, y).

components of the homogeneous region  $R_i$ , with  $(X_i^c, Y_i^c)$  the centroid of  $R_i$ .

If N(x,y) meets two or more of the labeled regions,  $\varphi(x,y)$  takes a value of *i* such that N(x,y)meets  $R_i$  and  $\delta(x,y,R_i)$  is minimized.

$$\varphi(x,y) = \min_{(x,y)\in H} \{\delta(x,y,R_j) | j \in \{1,\ldots,q\}\}$$
(3)

This seeded region growing procedure is repeated until all pixels in the image have been allocated to the corresponding regions. The definition of Eqs. (1) and (3) ensures that the final partition of the image is divided into a set of regions as homogeneous as possible on the basis of the given constraints. SRG algorithm is robust, rapid and free of tuning parameters and it is also very attractive for semantic image segmentation. However, SRG algorithm also suffers from the problems of pixel sorting and automatic seed selection.

#### 3. Automatic seeded region growing

An advantage of SRG is that the high-level knowledge of semantic image components can be exploited by selecting the suitable seeds for growing more meaningful regions. This property is very attractive for content-based image database applications (Fan et al., 2001). The natural questions which arise from a demonstration of SRG are: how to manage the pixel labeling procedure more efficiently? how to select the seeds automatically? how critical is the seed selection to a good segmentation? A poor starting estimate of region seeds or bad pixel sorting orders may result in an incorrect segmentation of an image (Mehnert and Jackway, 1997; Fan et al., 2001a). The obvious way to improve the SRG technique is to provide a more effective pixel labeling technique and automate the process of seed selection. In this section, we propose three automatic seed selection algorithms and an interesting boundary-oriented pixel labeling technique. We also give the performance comparison of these three automatic SRG techniques on a relatively large variety of images.

# 3.1. SRG via regular seed generation

The images can be first partitioned into a set of rectangular regions with fixed size as shown in Fig. 3. A simple automatic SRG algorithm can then be realized by selecting the centers of these rectangular regions as the seeds. However, the traditional sequential pixel sorting technique may meet the unconnected problem at the beginning of seeded region growing procedure. As shown in Fig. 4, it is hard to allocate the unlabeled pixel (x, y) to

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	9	0	0	0	0
0	0	0	6 <	Dg (	6	0	0	0	0
0	0	0	9	$\left  \left( \phi \right) \right $	Ø	0	0	0	0
0	0	0	م	0	0	0	0	0	0
0	6	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Fig. 3. The space partition of an image and the regular seed selection scheme.

unlabelled pixel



Fig. 4. The example for the unlabeled pixel at the beginning of seeded region growing.

any of the selected seeds because the pixel (x, y) is not adjacent to all the seeds. It is more reasonable to start the pixel labeling procedure from the selected seeds by involving new pixels step by step. Based on this observation, we propose a boundary-oriented pixel labeling technique to accelerate the SRG procedure, where the region growing is realized by dilating its boundaries step by step. Moreover, this boundary-oriented pixel labeling technique can support parallel region growing according to the following steps:

- The pixels in the same region are labeled by the same symbol. The adjacent regions connect via their common boundaries. The regions are represented via two parameters: one is the *centroid* of the region, and the other is a set of *boundary pixels*. These two region description parameters are updated by involving new pixels step by step. At the beginning of SRG procedure, the data sets for the centroid and the boundary pixels of a region are the same, i.e., the seed of the corresponding region.
- The SRG procedure starts from all the seeds at the same time. For a seeded region  $R_i$  with the set of boundary pixels  $B_{R_i} = \{(x_l, y_l) | l \in [1, ..., L]\}$ , we can test the second-order neighboring pixels  $(x_l \pm 1, y_l \pm 1)$  of its boundary pixel  $(x_l, y_l)$ . If the unlabeled pixel at  $(x_l \pm 1, y_l \pm 1)$  is similar with the adjacent boundary pixel  $(x_l, y_l)$  of the region  $R_i$ , then they are merged into the region  $R_i$  and also replace

the boundary pixel  $(x_i, y_i)$  as the new boundary pixel of the region  $R_i$ . This similarity testing and labeling procedure for all the boundary pixels for the same region can also be performed at the same time.

• The color similarity distance  $D(x_l, y_l, R_i)$ , between the unlabeled pixel  $(x_l \pm 1, y_l \pm 1)$  and the current testing boundary pixel  $(x_l, y_l)$  of the region  $R_i$ , is calculated as

$$D(x_{l}, y_{l}, R_{i}) = |I(x_{l}, y_{l}) - I(x_{l} \pm 1, y_{l} \pm 1)| + |u(x_{l}, y_{l}) - u(x_{l} \pm 1, y_{l} \pm 1)| + |v(x_{l}, y_{l}) - v(x_{l} \pm 1, y_{l} \pm 1)|$$

$$(4)$$

where  $I(x_l, y_l)$ ,  $u(x_l, y_l)$ , and  $v(x_l, y_l)$  indicate the values of the three color components of the testing boundary pixel  $(x_l, y_l)$ .  $I(x_l \pm 1, y_l \pm 1)$ ,  $u(x_l \pm 1, y_l \pm 1)$ , and  $v(x_l \pm 1, y_l \pm 1)$  represent the values of three color components of the unlabeled pixel  $(x_l \pm 1, y_l \pm 1)$  which is adjacent to the boundary pixel  $(x_l, y_l)$ . If the unlabeled pixel  $(x_l \pm 1, y_l \pm 1)$  meets two or more of the labeled boundary pixels of the region  $R_i$ , it is merged into the region  $R_i$  and replaces the most similar boundary pixel  $(x_m, y_m)$  of the region  $R_i$ :

$$D(x_m, y_m, R_i) = \min_{(x_l, y_l) \in B_{R_i}} \{ D(x_l, y_l, R_i) | l \\ \in \{1, \dots, L\} \}$$
(5)

• If the unlabeled pixel meets two or more boundary pixels from adjacent regions, it is merged into region *R<sub>j</sub>* which has the smallest similarity distance and replace the most similar boundary pixel as the new boundary pixel of the region *R<sub>j</sub>*:

$$D(x_k, y_k, R_j) = \min_{i \in \{1, \dots, q\}} \{ D(x_m, y_m, R_i) | (x_m, y_m) \in B_{r_i} \}$$
(6)

• The parallel SRG procedure for each boundary pixel will stop if the boundary pixels for the neighboring regions are connected or the color similarity distance is above a predefined threshold.

Since the seeds are selected regularly without involving the spatial distribution of image components, small image regions, whose sizes are less than the initial rectangular box, may be lost if the centers of the corresponding rectangular regions do not locate inside these small image regions. On the other hand, oversegmentation for the image regions with large size is also included because several seeds may be used for the same large region. In order to solve the first problem, new seeds are added automatically if the small regions can not be merged into their connected seed regions. In order to solve the second (oversegmentation) problem, a post-procedure of similarity-based region merging is performed on these adjacent regions.

In order to accelerate this similarity-based region merging procedure, the region adjacency graph (RAG) is built to express the relations among the regions (Tremeau and Colantoni, 2000). Each link of the RAG has a binary state: *on* or *off*, which determines the membership of a node to a given processing neighborhood. The RAG for an image is stored in a table and this table is updated along with the neighboring relations if the regions are merged (see Fig. 5). The region merging procedure is performed on the adjacent regions according to their color similarity. The color information of each image region is characterized by its color histogram (Swain and Ballard, 1991). The color similarity distance between two neighboring regions  $R_i$  and  $R_j$  can be defined as

$$d(R_i, R_j) = \sum_{l=0}^{N} \sum_{m=0}^{N} a_{lm} |H_i(l) - H_j(m)| \times |H_i(m) - H_j(l)|$$
(7)

where  $a_{lm}$  is a symmetric matrix of weights between 0 and 1 representing the similarity between bins l and m,  $H_i(l)$  and  $H_j(l)$  denote the color histogram bins for the lth color component, N is the maximum number of bins in the color histogram. The RAG for image regions is stored as a table in our experiments, and the color similarity distance of any two neighboring regions are calculated and stored in a distance table. If the color similarity distance  $d(R_i, R_j)$  is less than a predefined threshold, the adjacent regions  $R_i$  and  $R_j$  are merged. The color histogram of new region is then calculated, the similarity distance table and the RAG table are updated as shown in Fig. 5. This



Fig. 5. The table of region adjacency graph for region merging: (a) adjacency region graph before merging; (b) adjacency region graph after merging; (c) relationship table before region merging; (d) relationship table after region merging.

region merging procedure stops when all the color similarity distances in the table are above the predefined threshold.

For evaluating the real performance of this automatic SRG algorithm, we have tested a relatively large variety of color images. Fig. 6(a) shows an original image of "Akiyo", Fig. 6(b) is the automatic SRG result via regular seed selection. One can find that the image is oversegmented because the large homogeneous regions may span over several rectangular boxes and several seeds are used. Fig. 6(c) shows the SRG result after the post-procedure of similarity-based region merging, where the similar adjacent regions as shown in Fig. 6(b)are merged to form more meaningful image regions. The boundaries of these small regions (initial segmentation results) are also shown in Fig. 6(c) with low grey level, and this will help us to understand what kind of small regions are merged to form meaningful large regions.

# 3.2. SRG via edge-oriented seed generation

The simplified geometric structures of the image regions can be obtained from their color edges. An automatic edge-oriented seed generation technique has been proposed in (Fan et al., 2001a), where the initial seeds for SRG are obtained automatically from the *centroid* of the color edges. The centroid  $(X_{i,j}^{c}, Y_{i,j}^{c})$  between two adjacent labeled edge regions  $E_i$  and  $E_j$ , which is defined as the algebraic average of the edge pixels in the corresponding regions, is calculated as

$$\begin{cases} X_{i,j}^{c} = \frac{\sum_{(x,y) \in \{E_{i}, E_{j}\}} x}{\sum_{(x,y) \in \{E_{i}, E_{j}\}} \delta(x, y)} \\ Y_{i,j}^{c} = \frac{\sum_{(x,y) \in \{E_{i}, E_{j}\}} y}{\sum_{(x,y) \in \{E_{i}, E_{j}\}} \delta(x, y)} \end{cases}$$
(8)

where  $\delta(x, y) = 1$  if and only if  $(x, y) \in \{E_i, E_j\}$ , otherwise,  $\delta(x, y) = 0$ .

The boundary for the same homogeneous region may be partitioned into several adjacent edge regions because the obtained color edges are normally discontinuous, thus the centroids between several adjacent edge regions may be very close and their colors are also very similar. Therefore, these neighboring similar centroids are merged to one. These refined centroids are then taken as the initial seeds  $S_1, S_2, \ldots, S_q$ , for seeded region growing and these seeds are updated step by step by involving the new points. The color similarity



Fig. 6. The performance comparison of three automatic seeded region growing algorithm on "Akiyo": (a) original image; (b) region boundaries obtained by regular seed selection; (c) region boundaries after region merging, the boundaries for the small regions are also shown in low grey levels; (d) color edges; (e) region boundaries obtained by edge-oriented seed selection; (f) region boundaries obtained by the improved edge-oriented seed generation technique.

distance between the current testing pixel at (x, y)and its connected region  $R_i$  can be calculated via Eq. (4).

After the region seeds are obtained, the boundary-oriented pixel labeling technique as described in Section 3.1 is used for managing the similaritybased SRG procedure. The SRG procedure can start from all the seeds at the same time, and it stops if the boundary pixels for the neighboring regions are connected or the color similarity distance is above a predefined threshold. Fig. 6(d) shows the detected image edges. We can find that the simplified geometric structures of the images are given by the color edges even they are discontinuous. By using the centroids of these adjacent edges as the seeds, the SRG result is given in Fig. 6(e). One can find that the edge-oriented SRG technique can provide more meaningful segmentation results as compared with the SRG technique via regular seed selection, because the distribution of image components (provided by the color edges) are used for seed selection. Since the color edges are overdetected for the texture images, the segmentation results obtained by the edge-oriented SRG technique are not much better than those obtained by using the SRG technique via regular seed selection.

# 3.3. SRG via improved edge-oriented seed generation

The images may be corrupted by additional noise or they are textured. The edge detection algorithm may induce over-detections on the color edges. The edge-oriented seed generation technique, that takes the centroids of these connected color edges as the seeds, may induce the redundant seeds and thus results in oversegmentation of the images. There are three potential approaches to solving this oversegmentation problem: the first one is to perform a post-procedure of region merging as described in Section 3.1, the second one is to perform an image filtering procedure to reduce the noise and smooth the texture regions before the color edge detection is performed, the third one is to perform an edge smoothing and thinning procedure as done in the traditional boundary-based segmentation techniques (Kass et al., 1987).

In our current works, we use the second approach via image filtering. The window of image filter can be defined as

$$N_{m \times n}(x, y) = \left\{ (x + i, y + j)|_{-[m/2] \le i \le [m/2], -[n/2] \le j \le [n/2]} \right\}$$
(9)

where m, n are odd, [m/2] and [n/2] denote the largest integer not greater than the argument of m and n,  $N_{m \times n}(x, y)$  denotes a m-row and n-column image region. The image filtering procedure (image smoothing) tries to find a best trade-off between the noise reduction and the loss of image details. The general form of this processing is

$$\overline{I}(x,y) = \frac{1}{mn} \sum_{i=-[m/2]}^{[m/2]} \sum_{j=-[n/2]}^{[n/2]} \alpha_{x+i,y+j} I(x+i,y+j)$$
(10)

where  $\alpha_{x+i,y+j}$  is the weighting coefficient, I(x + i, y + j) indicates the grey level value of the pixel at (x + i, y + j), and  $\overline{I}(x, y)$  is the average grey level value of the pixel at (x, y) after the filtering procedure. The grey level value of pixel at (x, y)is then replaced by its corresponding average grey level value  $\overline{I}(x, y)$ . The small window size can be used to reduce the white Gaussian noise for color edge detection, while large window size is useful for detecting texture boundaries. Multiple scales can be used for solving the problem of texture image filtering (Bouman and Shapiro, 1994). In order to avoid the estimation of texture model parameter, the authors in (Deng and Manjunath, 2001) have proposed a color quantization technique to smooth the image colors into several representative classes. This color quantization technique can also be selected for image filtering by quantizing a set of image pixels to the same color. The same image filtering procedure is also performed on the other two chrominance components.

The color edge detection procedure is then performed on the smoothed images. The centroids of the neighboring labeled edges are taken as the seeds for automatic SRG as described in Section 3.2. Fig. 6(f) shows the segmentation result obtained by this improved edge-oriented SRG technique. One can find that its performance on simple image like "Akiyo" is very similar with the traditional edge-oriented SRG technique.

We have also tested a relatively large variety of color images from *Corel* database, and parts of the results are shown in Figs. 7 and 8. We also find that our improved edge-oriented SRG technique is very attractive for medical image segmentation, and parts of test results are given in Figs. 9 and 10.

# 4. Moving object extraction via seed tracking

During the last decade, many approaches to automatic moving object extraction have been proposed. The existing moving object extraction techniques can be classified into three categories:

• Temporal segmentation: Temporal segmentation only use motion information deduced from consecutive frames. A classical approach first consists in estimating a dense motion field and then partition the scene only based on the obtained motion information, where the adjacent video components are merged to form the meaningful video objects if they obey the same Hough or affine transformation motion model (Adiv, 1985; Wang and Adelson, 1994). However, dense field motion vectors are not very reliable for noisy data (Deng and Manjunath, 2001). In order to avoid the noise of optical flow, some techniques first include a change detector (Fan et al., 2001b; Diehl, 1991). However, the change detector induces the holes in the uniform regions.

- Spatiotemporal segmentation: To make the boundaries of moving objects correspond accurately to their spatial feature variation, spatiotemporal video segmentation techniques are proposed for moving object extraction via Bayesian frameworks or Markov random fields (Bouthemy and Francois, 1993; Moscheni et al., 1998; Gu et al., 1996; Alatan et al., 1998). However, their major drawbacks are the computational complexity and also the number of objects to be found has to be specified.
- *Temporal tracking*: Since the spatial segmentation can provide the accurate boundaries of the video objects regarding to color or texture, spatial segmentation can be integrated with a temporal tracking procedure for moving object extraction (Meier and Ngan, 1998). Moreover, the semantic information can be involved in the spatial segmentation procedure via a human computer interaction or object seed detection procedure (Gu and Lee, 1998; Fan and Elmagarmid, 2001). This temporal tracking approach is very attractive for content-based video database applications because the object extraction for this case can be performed on off-line.

In this section, we propose an automatic moving object extraction technique via seed tracking.



Fig. 7. The segmentation results obtained by the improved edge-oriented SRG technique: original image, color edges, region of object.



Fig. 8. The segmentation results obtained by the improved edge-oriented SRG technique: original image, color edges, region of object.



Fig. 9. The segmentation results obtained by the improved edge-oriented SRG technique: original image, color edges, region of object.

In order to detect the moving objects in a video sequence, two critical parameters should be used: the spatial information of the region boundaries of moving objects, and the temporal relationship information of the moving objects. Therefore, the temporal intensity changes can be used for selecting the suitable seeds for generating the regions which correspond to moving objects. These seeds, that are selected for moving object detection, should be part of these potential seeds that are obtained by the spatial segmentation procedure as described in Section 3. For this seed tracking technique to work, camera motion should be small with respect to the object motion, otherwise, camera motion compensation should be performed before the temporal change detection procedure (Fan and Elmagarmid, 2001). In this paper, we focus on moving object extraction from the



Fig. 10. The segmentation results obtained by the improved edge-oriented SRG technique: original image, color edges without filtering, color edges with filtering, region of object.

image sequences with a relatively small camera motion, thus the camera motion compensation is not performed in the work reported here. Our seed tracking technique takes the following steps for automatic moving object extraction:

- Spatial segmentation: The scene cut detection technique in (Fan et al., 2000) is first performed on a video sequence to obtain the video shots. The automatic SRG technique as described in Section 3.3 is only performed on the first frame of each video shot to obtain the boundaries of the object regions. In order to track these object regions among frames within a video shot, the automatic seed selection procedure as described in Section 3.3 is performed on each frame to extract all the potential seeds for generating the regions of moving objects.
- Change detection: Given two successive frames  $F_{t-1}$  and  $F_t$  of a video sequence, the motion measure between two coordinate pixels can be calculated as

where N(x, y) indicates the second-order neighborhood of the pixel at (x, y) as shown in Fig. 1, the constant C = 1 is used to avoid numerical instabilities. For reducing the influence of stationary texture background on intensity changes among frames, the motion measure feature FCON is normalized by the spatial intensity derivation. The temporal relationships between the coordinate pixels can then be classified into two opposite classes according to their motion strength: *changed pixel* versus *stationary pixel*.

$$\begin{cases} FCON(x, y) \ge \overline{T}, & \text{changed pixel} \\ FCON(x, y) < \overline{T}, & \text{stationary pixel} \end{cases}$$
(12)

where the threshold  $\overline{T}$  is determined automatically by using 1D entropic thresholding technique as introduced in (Fan et al., 2001a; Cheng et al., 2000). Fig. 11(a) shows a reference image of "Akiyo" and Fig. 11(b) is its spatial segmentation result obtained by the improved

$$FCON(x,y) = \frac{|I(x,y,t-1) - I(x,y,t)|^2 \sum_{(i,j) \in N(x,y)} |I(x,y,t) - I(x+i,y+j,t)|}{\sum_{(i,j) \in N(x,y)} |I(x,y,t) - I(x+i,y+j,t)|^2 + C}$$
(11)



Fig. 11. The temporal segmentation results for "Akiyo": (a) reference image; (b) regions for reference image; (c) original test image; (d) spatial segmentation regions for test image; (e) change pixels between test image and its reference; (f) moving object regions as compared with the spatial regions, the boundaries for the temporal unchanged regions are also shown in low grey levels.

edge-oriented SRG technique. Fig. 11(c) shows a test image of "Akiyo" and Fig. 11(d) is its spatial segmentation result obtained by our improved SRG technique. By performing the temporal change detection, Fig. 11(e) gives the temporal change pixels between the reference frame shown in Fig. 11(a) and the current test frame shown in Fig. 11(c). We also provide the spatial segmentation results to show how many regions are changed among frames.

• Seed tracking: The improved edge-oriented seed generation technique (without SRG procedure) as described in Section 3.3 is then performed on the current frame to obtain all the potential seeds. We are only interested in the moving objects (or regions of moving objects) and the moving objects should be located inside the temporal changed regions. The temporal change detection information are then used for selecting the suitable seeds from all the potential seeds. Only the seeds, which are located in the temporal changed regions, are selected for generating the moving regions of interest. If two connected regions are similar on color or texture, they are merged and considered to be the same moving object. Some moving regions, which are obtained by the colorbased region growing according to the selected seeds, may span over the temporal change mask. These regions are taken as the uncovered background because the moving objects should be fully located inside the temporal change mask. The new objects, which are fully located inside the temporal change mask like the moving objects, can also be detected by this seed tracking procedure. Fig. 11(f) shows the moving objects obtained by this seed tracking technique. We also visualize the spatial segmentation results in low grey levels to show which regions are changed among frames. One can find that these moving objects (eyes) are not the semantic objects that the video database users are concerned with.

The test results for "Salesman" and "Miss American" are given in Figs. 12 and 13. Since the camera motion is included in these sequences, the temporal change regions cover some redundant seeds (seeds for the moving background induced by camera motion). Camera motion compensation may first be performed for removing the effect of camera motion on the temporal changes.



Fig. 12. The temporal segmentation results for "Salesman": (a) reference image; (b) test image 1; (c) change pixels between test 1 and reference; (d) object regions for test 1, the boundaries for the temporal unchanged regions are also shown in low grey levels; (e) test image 2; (f) change pixels for test image 2; (g) object regions and uncovered background for test image 2, the boundaries for temporal unchanged regions are also shown in low grey levels.



Fig. 13. The temporal segmentation results for "Miss American": (a) reference image; (b) test image 1; (c) change pixels between test 1 and reference; (d) object regions for test 1, the boundaries for the temporal unchanged regions are also shown in low grey levels; (e) test image 2; (f) change pixels for test image 2; (g) object regions and uncovered background for test image 2, the boundaries for the temporal unchanged regions are also shown in low grey levels.

## 5. Salient object detection for image indexing

The salient objects are defined as the visually distinguishable image compounds. For example, the salient object "sky" is defined as the connected image regions with large sizes (i.e., dominant image regions) that are related to the human semantics "sky". We have already implemented 32 functions to detect 32 types of salient objects in natural scenes, and each function is able to detect a certain type of these salient objects in the basic vocabulary. Each detection function consists of three parts: (a) automatic image segmentation by using the techniques introduced in Section 3; (b) image region classification by using the SVM classifiers with an optimal model parameter search scheme and (c) labelbased region aggregation for automatic salient object generation.

We use our detection function of the salient object "sand field" as an example to show how we can design our detection functions. As shown in Fig. 14, image regions with homogeneous color or texture are first obtained by using the seeded region growing technique in Section 3. Since the visual properties of a certain type of salient object may look different at different lighting and capturing conditions, using only one image is insufficient to represent its visual characteristics. Thus this automatic image segmentation procedure is performed on a set of training images which consist of the salient object "sand field".

The homogeneous regions in the training images, that are related to the salient object "sand field", are selected and labeled as the training samples by human interaction. Region-based low-level visual features, such as 1-dimensional coverage ratio (i.e., density ratio) for a coarse shape representation, 6-dimensional region locations (i.e., 2dimensions for region center and 4-dimensions to indicate the rectangular box for a coarse shape representation), 7-dimensional LUV dominant colors and color variances, 14-dimensional Tamura texture, and 28-dimensional wavelet texture features, are extracted for characterizing the visual properties of these labeled image regions that are explicitly related to the salient object "sand field". The 6-dimensional region locations are used to determine the spatial contexts among different types of salient objects to avoid the wrong detection of the visual similar salient objects such as "beach sand" and "road sand".

We use *one-against-all* rule to label the training samples  $\Omega_{g_j} = \{X_l, L_j(X_l) | l = 1, ..., N\}$ : positive samples for the specific salient object "sand field" and negative samples. Each labeled training sample is a pair  $(X_l, L_j(X_l))$  that consists of a set of region-based low-level visual features  $X_l$  and the semantic label  $L_j(X_l)$  for the corresponding labeled homogeneous image region.

The image region classifier is learned from these available labeled training samples. We use the well-known SVM classifiers for binary image region classification (Vapnik, 1999). Consider a binary classification problem with linearly separable sample set  $\Omega_{g_i} = \{X_l, L_i(X_l) | l = 1, ..., N\}$ , where the semantic label  $L_i(X_l)$  for the labeled homogeneous image region with the visual feature  $X_l$  is either +1 or -1. For the positive samples  $X_l$  with  $L_i(X_l) = +1$ , there exists the transformation parameters A and b such that  $A \cdot X_l + b > +1$ . Similarly, for negative samples  $X_l$  with  $L_i(X_l) = -1$ , we have  $\Lambda \cdot X_l + b < -1$ . The margin between these two supporting planes will be  $2/||\Lambda||^2$ . The SVM classifier is then designed for maximizing the margin with the constraints  $A \cdot X_l + b > +1$  for the positive samples and  $\Lambda \cdot X_l + b < -1$  for the negative samples.

Given the training set  $\Omega_{g_j} = \{X_l, L_j(X_l) | l = 1, ..., N\}$ , the margin maximization procedure is then transformed into the following optimization problem:

$$\underset{A,b,\xi}{\operatorname{arg\,min}} \frac{1}{2} \Lambda^{\mathrm{T}} \cdot \Lambda + C \sum_{l=1}^{N} \xi_{l}$$

$$L_{j}(\Lambda \cdot \Phi(X_{l}) + b) \ge 1 - \xi_{l}$$
(13)

where  $\xi_l \ge 0$  represents the training error rate, C > 0 is the penalty parameter to adjust the training error rate and the regularization term  $\frac{\Delta^{T} \cdot A}{2}$ ,  $\Phi(X_l)$  is the function that maps  $X_l$  into higherdimensional space (i.e., feature dimensions plus the dimension of response) and the kernel function is defined as  $\kappa(X_i, X_j) = \Phi(X_i)^T \Phi(X_j)$ . In our current



Fig. 14. The flowchart for automatic salient object detection.

implementation, we select radial basis function (RBF),  $\kappa(X_i, X_j) = \exp(-\gamma ||X_i - X_j||^2), \gamma > 0.$ 

We have developed an efficient search algorithm to determine the optimal model parameters  $(C, \gamma)$ for the SVM classifiers: (a) The labeled image regions are partitioned into v subsets in equal size, where v - 1 subsets are used for classifier training and the remaining one is used for classifier validation. (b) Our feature set for image region representation is first normalized to avoid the features in greater numeric ranges that dominate those in smaller numeric ranges. Because inner product is usually used to calculate the kernel values, this normalization procedure is able to avoid the numerical problem. (c) The numeric ranges for the parameters C and  $\gamma$  are exponentially partitioned into small pieces with M pairs. For each pair, v - 1 subsets are used to train the classifier model. When the M classifier models are available, cross-validation is then used to determine the underlying optimal parameter pair  $(C, \gamma)$ . (d) Given the optimal parameter pair  $(C, \gamma)$ , the final classifier model (i.e., support vectors) is trained again by using the whole training data set. (e) The spatial contexts among different types of salient objects (i.e., coherence among different types of salient objects) have also been used to cope well with the wrong detection problem for the visual similar salient objects.

Some results for our detection functions are shown in Figs. 15 and 16. From these experimental results, one can find that the salient objects are more representative than the homogeneous image regions and the major visual properties of the dominant image components are maintained by using the salient objects for image content representation. Thus using the salient objects for feature extraction can enhance the quality of features and result in more



Fig. 15. The detection results of the salient object "water".



Fig. 16. The detection results of the salient object "sand field".

effective semantic image classification. In addition, the salient objects can be visually distinguishable, and they are also semantic to human beings. Thus the keywords for interpreting the salient objects can also be used to achieve the annotations of the images at the content level. The average performance for some detection functions is given in Table 1. It is worth noting that the procedure for salient object detection is automatic and the human interaction is only involved in the procedure to label the training samples (i.e., homogeneous image regions) for learning the detection functions.

# 6. Conclusion

Automatic image segmentation has become the key point for realizing content-based image description and retrieval, and object-based image

compression. SRG algorithm, which is robust, rapid and free of tuning parameters, is very attractive for semantic image segmentation. However, the traditional SRG algorithm suffers from the problems of pixel sorting orders for labeling and automatic seed selection. An automatic SRG algorithm, along with a more effective pixel labeling technique and an automatic seed selection method, is presented in this paper. A seed tracking algorithm is also proposed for automatic moving object extraction, where the seeds located inside the temporal change mask are selected for generating the regions of moving objects. Our seed tracking technique is very attractive for detecting the uncovered background and new objects which are also located inside the temporal change mask like the moving objects. Our future research will focus on how to handle the limitations in the algorithm, such as improving its performance for tex-

	Salient objects		
	Brown horse	Grass	Purple flower
ρ	95.6%	92.9%	96.1%
Q	100%	94.8%	95.2%
	Red flower	Rock	Sand field
ρ	87.8%	98.7%	98.8%
Q	86.4%	100%	96.6%
	Water	Human skin	Sky
ρ	86.7%	86.2%	87.6%
Q	89.5%	85.4%	94.5%
	Snow	Sunset/sunrise	Waterfall
ρ	86.7%	92.5%	88.5%
Q	87.5%	95.2%	87.1%
	Yellow flower	Forest	Sail cloth
ρ	87.4%	85.4%	96.3%
Q	89.3%	84.8%	94.9%
	Elephant	Cat	Zebra
ρ	85.3%	90.5%	87.2%
Q	88.7%	87.5%	85.4%

Table 1 The average performance of some detection functions (precision  $\rho$  versus recall  $\rho$ )

ture images and video sequences with a relatively large camera motion.

#### References

- Adams, R., Bischof, L., 1994. Seeded region growing. IEEE Trans. Pattern Anal. Machine Intell. 16, 641–647.
- Adiv, G., 1985. Determining three-dimensional motion and structure from optical flow generated by several moving objects. IEEE Trans. Pattern Anal. Machine Intell. 7, 384– 401.
- Alatan, A.A., Onural, L., Wollborn, M., Mech, R., Tuncel, E., Sikora, T., 1998. Image sequence analysis for emerging interactive multimedia services—The European COST 211 framework. IEEE Trans. Circ. Syst. Vid. Technol. 8, 802– 813.
- Bouman, C.A., Shapiro, M., 1994. A multiscale random field model for Bayesian image segmentation. IEEE Trans. Image Process. 3, 162–177.
- Bouthemy, P., Francois, E., 1993. Motion segmentation and qualitative dynamic scene analysis from an image sequence. Internat. J. Comput. Vision 10, 157–182.
- Chang, Y.-L., Li, X., 1994. Adaptive image region-growing. IEEE Trans. Image Process. 3, 868–872.

- Cheng, H.D., Chen, Y.H., Jiang, X.H., 2000. Thresholding using two-dimensional histogram and fuzzy entropy principle. IEEE Trans. Image Process. 9, 732–735.
- Chu, C., Aggarwal, J.K., 1993. The integration of image segmentation maps using region and edge information. IEEE Trans. Pattern Anal. Machine Intell. 15, 1241– 1252.
- Deng, Y., Manjunath, B.S., 2001. Unsupervised segmentation of color-texture regions in image and video. IEEE Trans. Pattern Anal. Machine Intell. 23 (8), 800–810.
- Diehl, N., 1991. Object-oriented motion estimation and segmentation in image sequence. Signal Process.: Image Comm. 3, 23–56.
- Fan, J., Elmagarmid, A.K., 2001. An automatic algorithm for semantic object generation and temporal tracking. Signal Process.: Image Comm. 16.
- Fan, J., Yau, D.K.Y., Aref, W.G., Rezgui, A., 2000. Adaptive motion-compensated video coding scheme toward contentbased bit rate allocation. J. Eletrical Imaging 9, 521–533.
- Fan, J., Yau, D.K.Y., Elmagarmid, A.K., Aref, W.G., 2001a. Automatic image segmentation by integrating color-based extraction and seeded region growing. IEEE Trans. Image Process. 10 (10), 1454–1466.
- Fan, J., Yu, J., Fujita, G., Onoye, T., Wu, L., Shirakawa, I., 2001b. Spatiotemporal segmentation for compact video representation. Signal Process.: Image Comm. 16, 553–566.

- Grinias, I., Tziritas, G., 2001. A semi-automatic seeded region growing algorithm for video object localization and tracking. Signal Process.: Image Comm. 16, 977–986.
- Gu, C., Lee, M.C., 1998. Semantic segmentation and tracking of semantic video objects. IEEE Trans. Circ. Syst. Vid. Technol. 8, 572–584.
- Gu, H., Shirai, Y., Asada, M., 1996. MDL-based segmentation and motion modeling in a long image sequence of scene with multiple independently moving object. IEEE Trans. Pattern Anal. Machine Intell. 18, 58–64.
- Haddon, J., Boyce, J., 1990. Image segmentation by unifying region and boundary information. IEEE Trans. Pattern Anal. Machine Intell. 12, 929–948.
- Haralick, R.M., Shapiro, L.G., 1985. Survey: Image segmentation techniques. Comput. Vision Graphics Image Process. 29, 100–132.
- Haris, K., Efstratiadis, S.N., Maglaveras, N., Katsaggelos, A.K., 1998. Hybrid image segmentation using watersheds and fast region merging. IEEE Trans. Image Process. 7 (12), 1684–1699.
- Hijjatoleslami, S.A., Kittler, J., 1998. Region growing: A new approach. IEEE Trans. Image Process. 7, 1079–1084.
- Kass, M., Witkin, A., Terzopoulos, D., 1987. Snakes: Active contour models. In: Proc. 1st Internat. Conf. on Computer Vision, pp. 259–267.
- Kunt, M., Benard, M., Leonardi, R., 1987. Recent results in high-compression image coding. IEEE Trans. Circ. Syst. 34, 1306–1336.
- Lim, Y.W., Lee, S.U., 1990. On the color image segmentation algorithm based on the thresholding and the fuzzy C-means technique. Pattern Recognition 23 (9), 935–952.
- Majunath, B.S., Huang, T.S., Tekalp, A.M., Zhang, H.J., 2000. special issue on image and video processing for digital libraries. IEEE Trans. Image Process. 9 (1).

- Meier, T., Ngan, K.N., 1998. Automatic segmentation of moving objects for video object plane generation. IEEE Trans. Circ. Syst. Vid. Technol. 8, 525–538.
- Mehnert, A., Jackway, P., 1997. An improved seeded region growing algorithm. Pattern Recognition Lett. 18, 1065– 1071.
- Moscheni, F., Bhattacharjee, S., Kunt, M., 1998. Spatiotemporal segmentation based on region merging. IEEE Trans. Pattern Anal. Machine Intell. 20, 897–914.
- Pal, N., Pal, S., 1993. A review on image segmentation techniques. Pattern Recognition 26, 1277–1294.
- Palmer, P.L., Dabis, H., Kittler, J., 1996. A performance measure for boundary detection algorithms. Computer Vision and Image Understanding 63, 476–494.
- Pappas, T.N., 1992. An adaptive clustering algorithm for image segmentation. IEEE Trans. Signal Process. 40, 901–914.
- Pavlidis, T., Liow, Y.T., 1990. Integrating region growing and edge detection. IEEE Trans. Pattern Anal. Machine Intell. 12, 225–233.
- Sahoo, P.K., Soltani, S., Wong, A.K.C., 1988. A survey of thresholding techniques. Comput. Vision Graphics Image Process. 41, 233–260.
- Shen, X., Spann, M., Nacken, P.F.M., 1998. Segmentation of 2D and 3D images through hierarchical clustering. Pattern Recognition 31, 1295–1320.
- Swain, M., Ballard, D., 1991. Color indexing. Internat. J. Comput. Vision 7 (1).
- Tremeau, A., Colantoni, P., 2000. Region adjacent graph applied to color image segmentation. IEEE Trans. Image Process. 9, 735–744.
- Vapnik, V., 1999. The Natural of Statistical Learning Theory. Springer.
- Wang, J., Adelson, E., 1994. Representing moving image with layers. IEEE Trans. Image Process. 3, 625–638.