Natural color image segmentation using integrated mechanism

Jie Xu (徐 杰) and Pengfei Shi (施鹏飞)

Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200030

Received March 12, 2003

A new method for natural color image segmentation using integrated mechanism is proposed in this paper. Edges are first detected in term of the high phase congruency in the gray-level image. K-mean cluster is used to label long edge lines based on the global color information to estimate roughly the distribution of objects in the image, while short ones are merged based on their positions and local color differences to eliminate the negative affection caused by texture or other trivial features in image. Region growing technique is employed to achieve final segmentation results. The proposed method unifies edges, whole and local color distributions, as well as spatial information to solve the natural image segmentation problem. The feasibility and effectiveness of this method have been demonstrated by various experiments.

OCIS code: 100.2960.

Unsupervised segmentation in general-purpose natural color images is a challenging problem because of variabilities of color, texture and shape in them. Many existing segmentation techniques that focus on gray-scale, color or texture images, usually do not work well on this kind of images. Only several techniques have been developed for natural image segmentation recently[1−5].

Region based image segmentation technique can provide closed region boundaries but it is quite hard to define region models which are sufficiently robust and general to allow "just enough" intra-region variation, without causing ambiguities in what constitutes a region. There is also a fundamental trade-off in edge detection between spatial localization and noise immunity as shown by Canny[6]. So it is clearly attractive to combine region processing with edge detection to get closed segmentation results.

Furthermore, the criteria by which we judge a segmentation is whether the result agrees with that performed by human vision system (HVS). Therefore, when we do segmentation, we should consider sufficiently the character of HVS. We have noticed that HSV (hue, saturation, value) has an ability to accommodate itself to whole color circumstance just as it can adjust itself to different light conditions. For this reason, two colors could be regarded as the same in one image, but as different in another. When we measure similarities of colors, we should take into consideration the global color circumstance. Moreover, HSV is not sensitive to color's difference in texture area. So the local spatial distribution of colors is also important. In this paper, we propose an approach of natural image segmentation, integrating information of edge, edge's position, local color difference and overall color distribution in image using an integrated mechanism.

Edges are first detected in term of high phase congruency in the gray-level image. Edge detection is one of the most widely studied problems in image analysis. Methods looking for maxima in the gradient of intensity, such as Canny edge operator[6], are inadequate for edges composed of combinations of steps, peaks and roofs, which are popular in natural color images. Morrone and Owens[7] proposed the phase congruency model for feature detection. Phase congruency is a dimensionless quantity that is invariant to change in image brightness or contrast. It also shows good localization and explains a number of psychophysical effects in human feature perception. A wide range of feature types, even Mach bands give rise to points of high phase congruency. However, it is a rather awkward quantity to calculate.

Kovesi proposed an approach to calculate phase congruency via log Gabor wavelets[8]. Wavelet transform was accomplished by image convolution with polar-separable Gaussian filters in the frequency domain. 1D log Gaussian in the radial direction has a transfer function of the form

$$g(w) = \exp \frac{-(\log (w/w_0))^2}{2(\log (k/w_0))^2}, \tag{1}$$

where w_0 is the filter's center frequency. k/w_0 keeps constant for varying w_0. The Gaussian cross-section in the angular direction is defined as

$$G(\theta) = \exp - \frac{(\theta - \theta_0)^2}{2\sigma_0^2}, \tag{2}$$

where θ_0 is the orientation angle of the filter, and σ_0 is the standard deviation of the Gaussian function in the angular direction. Then 2D phase congruency is computed as

$$E_{no}(x) = A_{no}(x)\Delta\phi_{no}(x), \tag{3}$$

$$PC_2(x) = \frac{\sum_o \sum_n W_0(x)[E_{no}(x) - T_0]}{\sum_o \sum_n A_{no}(x) + \varepsilon}, \tag{4}$$

where $A_{no}(x)$ is the amplitude of the filter pair at position x, o and n denote the index over orientation and scale, respectively. ε is used to avoid division by zero, and to discount the result, both $A_{no}(x)$ should be very small. The compensation of noise $T(x)$ is defined as

$$T = \mu_{\mathrm{R}} + k\sigma_{\mathrm{R}}, \tag{5}$$

where μ_{R} and σ_{R} are the mean and variance of Rayleigh distribution, respectively. The weighting function $W(x)$

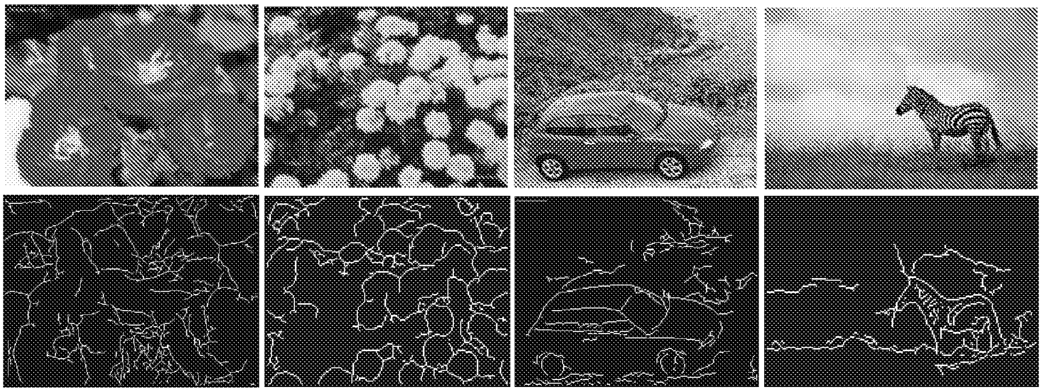

Fig. 1. Original natural images and the corresponding edge images.

is defined as

$$W(x) = \frac{1}{1 + e^{\gamma(c - s(x))}}, \tag{6}$$

where $s(x) = \frac{1}{N}\left(\frac{\sum_n A_n(x)}{\varepsilon + A_{\max}(x)}\right)$, N is the number of scales, $A_{\max}(x)$ is the amplitude of the filter pair having maximum response at position x. ε, c and γ are constants.

The phase deviation function is defined as

$$\Delta\Phi(x) = \cos(\phi_n(x) - \bar{\phi}(x)) - |\sin(\phi_n(x) - \bar{\phi}(x))|. \tag{7}$$

The edge maps are obtained by performing nonmaximal suppression and hysteresis thresholding on the raw phase congruency images.

In our experiments, we used two-octave bandwidth filters over four scales and six orientations to obtain local frequency information. The scaling between successive scales was 2. The ratio between the angular spacing of the filters and angular standard deviation of the Gaussian was 1.2. The other parameters needed in Eq. (4) was the same as in Ref. [8] (where $k = 2.0$, $c = 0.4$, $\gamma = 10$ and $\varepsilon = 0.01$). None of these parameter values is particularly critical. Figure 1 shows some examples for edge detection based on phase congruency. It should be emphasized that all the results were obtained by applying the same parameter and threshold values to every image.

We have provided simplified geometric structures of the image regions using the edge detector. However, these edges are normally discontinuous or over-detected. Then we use region growing (RG) technique to get closed segmentation results. Our strategy includes two stages. In the first stage, we cluster the long edge lines based on overall color information to estimate the rough distribution of regions, and merge the short ones based on local color and position information to eliminate the negative influence of local features.

The 8-neighbor edge pixels are connected into a line. For each line i, a descriptor is defined as

$$C_i = \{h, s, v, \text{percent}, cx, cy\}, \tag{8}$$

where h, s and v are the three components of the average color in the perceptually uniform CIE HSV color space. 'percent' is the ratio of the number of pixels on line i to the total number of pixels in image. (cx, cy) is the position of 'centroids', which is defined as the algebraic average of pixels and calculated as

$$cx = \bar{x} = \frac{1}{n}\sum_{(x,y)\in i} x, \quad cy = \bar{y} = \frac{1}{n}\sum_{(x,y)\in i} y. \tag{9}$$

Short edge lines usually reflect the local character in images, while long ones are more important to define a region with more possibility of being contours. We utilize K-mean clustering to class the long edge lines with a high level homogeneity to predict the approximate distribution of different objects. Long edge lines whose percentage is more than threshold Tl are first selected as independent clusters. The Euclidean distances on average colors of each pair among the clusters are calculated as

$$d_i^c(k,l) = \sqrt{(h_k - h_l)^2 + (s_k - s_l)^2 + (v_k - v_l)^2}. \tag{10}$$

Then we search for the shortest distance d_i and merge the two corresponding clusters into one, and the average color of the new cluster and the distance from it to every other cluster are updated. Repeat the clustering process until the difference Δd of the distances d_i and d_{i+1} in two successive process is greater than threshold T_d. A smaller threshold T_d is used to ensure the homogeneity within each cluster. Ignoring short edge lines, that are more sensitive to noise, also promote the purity within clusters. The edge lines in the same cluster are thought from the same object. Then we get the rough distribution of objects in the image according to the labelled lines. The short edge lines left unlabelled in the clustering process are considered as trivial details, which are usually caused by texture of object. The pair among them where the distance of centroids is nearest is always merged into one if their color difference is not exceeding the threshold T_c.

To this end, according to the labelled edge lines, we have attained the distribution of different objects in the image. In the second stage, RG is used to get segmentation results with the initial seeds generated from these labelled edge points automatically. With these suitable seeds, the high-level knowledge of semantic image components can be exploited for more meaningful region growing.

Each step of RG incorporates one additional pixel into

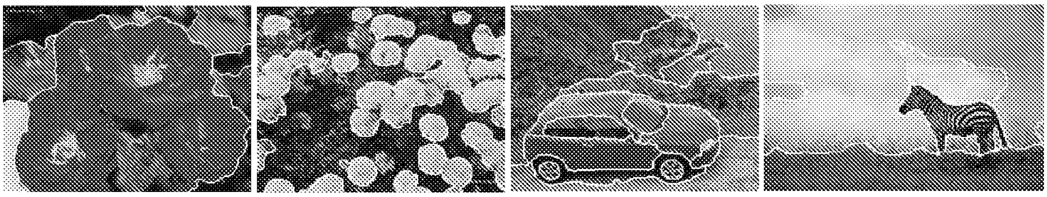

Fig. 2. Segmentation results.

one of the seeds. These initial seeds are further replaced by the pixels of the generated homogeneous regions. The pixels in the same region are labelled by the same symbol and by the different symbol in variant regions. Let $P(x,y)$ be an unallocated pixel which is adjacent to at least one of the labelled regions. If the distance between this pixel and its adjacent labelled region, which is calculated as Eq. (10), is less than threshold T_p, $P(x,y)$ joins this region. If $P(x,y)$ meets two or more of the labelled regions, it is allocated to the region with a minimal distance. The RG procedure is repeated until all the pixels in the image have been allocated to the corresponding regions. It should be noticed that all the short edge lines merged together would be treat as one pixel in this procedure. Thus, the regions formed will be robust to local varieties caused by the texture or trivial features.

The process of edge detection and RG ensure that the image is divided into a set of regions as homogeneous as possible based on the given constraints. This results in oversegmented regions because the initial given seeds for them are normally over detected. Hence, the adjacent regions, which do not have a significant color difference between them, are merged together to form a bigger region. The procedure is as follows. First, distances between the color histograms of any two neighboring regions, computed using the histogram intersection technique[9], are calculated and stored in a distance table, The pair of regions with the minimum distance is merged together. The color feature vector for the new region is calculated and the distance table is updated along with the neighboring relationships. The process continues until a maximum threshold T_r for the distance is reached. Moreover, any region that has fewer pixels than a certain percentage of the total number of image pixels is also merged into the most similar neighbor region.

We have applied this proposed method on a number of real natural images from Internet. These images are of JPG type, with the sizes of 192×128. Figure 2 shows some of segmentation results. The boundaries of different regions are highlighted using the white lines. According to the results, the proposed method shows great capabilities in distinction of different regions, while it is also robust to texture area. Not like other region growing method, regions obtained with our strategy, could be composed of several parts that may not be adjacent to each other. In general, these segmentation results match well with the perceived boundaries.

Many existing segmentation techniques, such as direct clustering methods in color space, only work well on homogeneous color regions. In this work, a natural image segmentation technique using integrated mechanism is proposed, which treated the interplay of colors and their spatial distribution in an inseparable way as they are actually perceived during the preattentive stage of human color vision. The proposed method is robust to texture and other trivial features in the image, and it also has shown stable results for different types of natural images.

J. Xu's e-mail address is may@sjtu.edu.cn.

References

1. W. Y. Ma and B. S. Manjunath, IEEE Trans. Image Processing **9**, 1375 (2000).
2. Y. Deng and B. S. Manjunath, IEEE Trans. Pattern Analysis and Machine Intelligence **23**, 800 (2001).
3. M. Mirmehdi and M. Petrou, IEEE Trans. Pattern Analysis and Machine Intelligence **22**, 142 (2000).
4. J. Fan, D. K. Y. Yau, A. K. Elmagarmid, and W. G. Aref, IEEE Trans. Image Processing **10**, 1454 (2001).
5. K.-M. Chen and S.-Y. Chen, Pattern Recognition Lett. **23**, 755 (2002).
6. J. Canny, IEEE Trans. Pattern Analysis and Machine Intelligence **8**, 679 (1986).
7. M. C. Morrone and R. A. Owens, Pattern Recognition Lett. **6**, 303 (1987).
8. P. Kovesi, Videre: J. Computer Vision Research **1** (3), 1 (1999).
9. M. J. Swain and D. H. Ballard, J. Computer Vision **7**, 11 (1991).