

Clustering Color Image Segmentation Based on Maximum Entropy

Sima Haifeng

School of Computer Science and Technology ,Henan
Polytechnic University
School of Computer Science and Technology , Beijing
Institute of Technology

Liu Lanlan

School of Emergence Management
Henan Polytechnic University

Abstract-Maximum entropy is meaningful for representing pixels spatial distribution in the image. This paper proposed a new clustering segmentation approach for color image according to the maximum entropy. Firstly, quantize the HSV color space to equal intervals. The probability distribution of pixels in the quantized space can be seen as a random process. Select a slide interval on the histogram to estimate the classes based on the maximum entropy in the color space. Then observed class number and the initial cluster center. Segmented pixels in to regions by clustering and used spatial filtering to eliminate meaningless regional and holes. The experiment results has shown that this algorithm achieved a good segmentation.

Keywords-Color image, Segmentation, Maximum entropy theory, K-means clustering

I. INTRODUCTION

In the computer vision field, image segmentation plays a crucial role as a preliminary step for high-level image processing. The study of gray image is much earlier, the algorithm is also relatively mature. In accompany of advances of relational technology, color images are applied widely. In recent years color image segmentation has aroused more and more attention, Compared with the gray image, color image includes not only the brightness information, but also more effective information: hue and saturation. In the same scene, there contained much more information in color images than the gray image. Color image is more sensitive to human perception. A low-quality color image seems more attractive than a perfect grayscale images. Therefore, the segmentation of color images is a broader research field.

Image segmentation is the basis of image analysis and understanding. The common approach including edge detection , spatial clustering and region growing method. In the analysis of the relationship between vector quantization and clustering based on the minimum sum of squares as a criterion, Michael,et,al [1] used competitive learning optimal weight vector as the cluster center, to achieve a color image segmentation. Lim,et,al [2]combined the FCM and histogram threshold for color image segmentation. The segmentation was divided into two stages. This approach attempts to determine the cluster number and center in the crude stages, and solve the problem of large computation. However, it has the same shortcomings as most of the histogram method.

Jayanta Mukherjee[3] proposed a general framework for refining initial clusters in a feature space using MRF processing is presented and subsequently, the algorithm for MRF processing in the color spaces is proposed. The proposed MRF processing is shown to be working with the principles of 1-NN and KNN classification rules among the neighbors of a pixel .

K means clustering algorithm [4] is the most popular cluster analysis methods and widely used. K means clustering algorithm for image segmentation is intuitive, fast and easy to implement features, but the biggest problem is to determine the number of clusters. Initial cluster centers impact on the classification results greatly. On the other hand, clustering algorithm use the color information of pixels to segment images, but do not consider the spatial information. Therefore, it is sensitive to noise and result in over-segmentation. Jong Hyun Park[5] propose an unsupervised segmentation algorithm for color images based on Gaussian mixture models (GMMs). The number of mixture components is determined automatically by adaptive mean shift, in which local clusters are estimated by repeatedly searching for higher density points in feature vector space. For the estimation of parameters of GMMs, the mean field annealing expectation-maximization (EM) is employed.

Thresholding methods [6, 7, 8] work well on the images with obvious histogram peak. The essence of segmentation is pixel classification. The class number determination is key step for segmentation of multi-objective image. This paper proposed a new clustering segmentation approach for color image according to the maximum entropy. Firstly, quantize the HSV color space to equal intervals. The probability distribution of pixels in the quantized space can be seen as a random process. Select a slid size interval on the hisgram , estimates the classes based on the maximum entropy in the color space and observed class number and the initial cluster center. Segmented pixels in to regions by clustering and used spatial filtering to eliminate meaningless regional and holes. The experiment results has shown that this algorithm achieved a good segmentation.

II. MAXIMUM ENTROPY FOR SEGMENTATION

In the information theory, entropy is a concept used to measure the amount of information. It is defined as: the average self-information is the order degree overall system. The more ordered a system, entropy becomes; the contrary, the more chaotic a system, the higher the entropy. Therefore,

the information entropy of the system can also be said to be a measure of the degree of ordering.

In the traditional binary segmentation of grayscale image, it chooses a single threshold to determine the foreground and background regions. This method of selecting threshold value is proved to be more effective than of OTSU and FCM segmentation on threshold selection method. For a color image segmentation in the color space, choose proper threshold for segmentation can appeal to maximum entropy. Such as color image in RGB space, it is clear there are three characteristics of each pixel under, how to select a reasonable threshold of three channels is a problem. It is very difficult to find the relevance between three elements in the RGB space and the meaning of the data for the segmentation. So we try to divide the color space into reasonable bins, calculate the statistical distribution of all pixels. we select threshold by maximum entropy on dynamic range of a regional partition. Partition of the color space is helpful to distinguish areas in the image. The three channels of the hsv color space are described as: hue[0,360°], saturation [0,1], value [0,1]. Divide the Hue space into 12 intervals; Saturation and Value were divided into 5 intervals. So that HSV space is divided into 300 bins. So the information source of pixels in bin can be define as : $X=[x_1, x_2, \dots, x_i, \dots, x_{300}]$, its probability space is $P=[P_1, P_2, \dots, P_i, \dots, P_{300}]$.

The essence of segmentation is pixels classification of image. The number of classes determines multi-objective image recognition. We define the sliding window on the X axis. The window size can be adjusted, then the number of pixels in a particular region to achieve the maximum overall when he was set to a class. Find the second largest value until the number of consecutive rest does not meet the window size, or number of pixels smaller than a certain threshold. Stop for regional estimates. At this time we got an image of the preliminary estimate of the number of clusters. Compute the pixel distribution of local entropy between interval [a,b], the maximum extent reflects the image of a region, and contains the max image information. The maximum value of a region is computed as follows :

$$H(R) = \int_a^b -p(x) \log p(x) \quad (1)$$

Adopt K-means clustering to segment pixels to proper regions. Initial K(c) value as estimated from above method. The hue, saturation and value values of the center intervals are defined as the initial clustering centers. The clustering algorithm is described as follows :

A Set estimated class centers as the initial cluster center sets K (c);

B For any pixels, find its distance to the c centers, defined the samples to the center of the shortest special distance ,the distance can be computed as:

$$\|X - C_i\| = \sqrt{(v - v_c)^2 + (s - s_c)^2 - 2 \cos(h - h_c)} \quad (2)$$

C Computing the distances between any two classes. When the between-class distances are smaller than the maximum distance within the class, combined the class, and update the class number and class centers until there is no

changed of classes. Update the center of the class by mean value. Compute the new H (new) of segmented image to ensure the clustering to be the most suitable to segment the image. When $H(\text{new}) > H(i)$, update the centers of all the clusters, else stop the iteration. Then reach the global maximum value of image.

$$H(i) = \sum_{i=1}^k H(R_i) \quad (3)$$

D For all cluster centers, while the k centers are unchanged after iteration steps (2) (3), the algorithm is over.

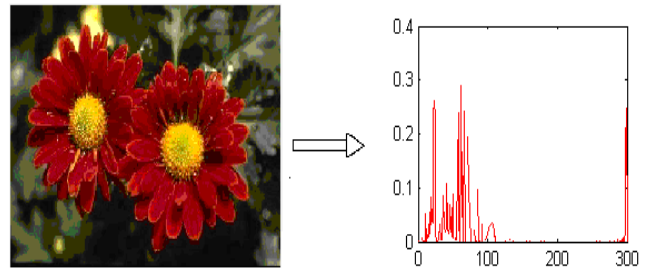


Figure 1. Color image and entropy of bins

III. FILTER DESIGN

Eliminate empty area and merge homogeneous regions. Spatial clustering of pixels classification will bring small parts of the inconsistencies and impurities. In order to reduce this impact on the overall division, use spatial filtering to eliminate the noise, segmentation results more complete and consistent visual experience.

A statistic class label of pixels in neighborhood within radius r; set different weight by formula 4) to calculate the impact on he center pixel. The weight function is shown in figure 2. Take the the corresponding class label with maximum value as the class of the pixel.

$$W(i) = e^{-r^3} \quad (4)$$

B detect small areas in the image, the proportion of local pixels less than the threshold value 0.5%, merge the small region into similar neighboring areas.

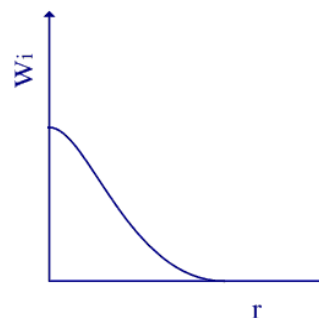


Figure 2. Weights of neighbor pixels

In order to assess the performance of the proposed algorithm, we have conducted experiments on the natural color images. The experiments are performed on over 50 images selected from Berkeley segmentation database and

other sources. Figure 3. illustrates some of the images have been segmented. From the results (a-f), we can see that the results observe accurate segmented regions in the images. The segmented regions match with color distribution in the original image well and the important details are preserved in the images.

IV. SUMMARIES

This paper has presented an unsupervised segmentation for natural color images. An effective segmentation method was proposed using entropy maximum estimation for clustering. The entropy maximum estimation has been used to determine the number of clusters in K-mean for the pixels of image. Experimental results of natural color images show that the estimation of clusters and segmentation performs well in HSV color feature spaces. The experimental results have shown the superiority of the algorithm. The proposed approach can be used in many image processing applications.

REFERENCES

- [1] Michael, T, U, Arbib, A. Color image segmentation using competitive learning [J], IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(12), PP: 1197–1206.
- [2] Lim Y W, Lee SU. On the color image segmentation algorithm based on the thresholding and the fuzzy c-mean s techniques [J]. Pattern Recognition, 1990, 23 (9), PP: 935–952.
- [3] Jayanta Mukherjee. MRF clustering for segmentation of color images. Pattern Recognition Letters 23 (2002), PP: 917–929.
- [4] CHEN TW, CHEN YL, CH IEN SY. Fast image segmentation based on K2means clustering with histograms in HSV color space[C]// MMSP 2008: Image /Video Processing and Coding. [S.l.]: IEEE, 2008, PP: 322–325.
- [5] Jong Hyun Park, Guee Sang Lee, Soon Young Park. Color image segmentation using adaptive mean shift and statistical model-based methods. Computers and Mathematics with Applications. 57 (2009), PP: 970–980.
- [6] Yen J C, Chang F J, Chang S. New criterion for automat c multilevel thresholding [J]. IEEE Transaction on Image Processing, 1995, 4(3), PP:370–378.
- [7] Kurugollu F, Sankur B, HarmanciA E. Color image segmentation using histogram multi thresholding and fusion [J]. Image and Vision Computing, 2001,19 (13), PP:915–928.

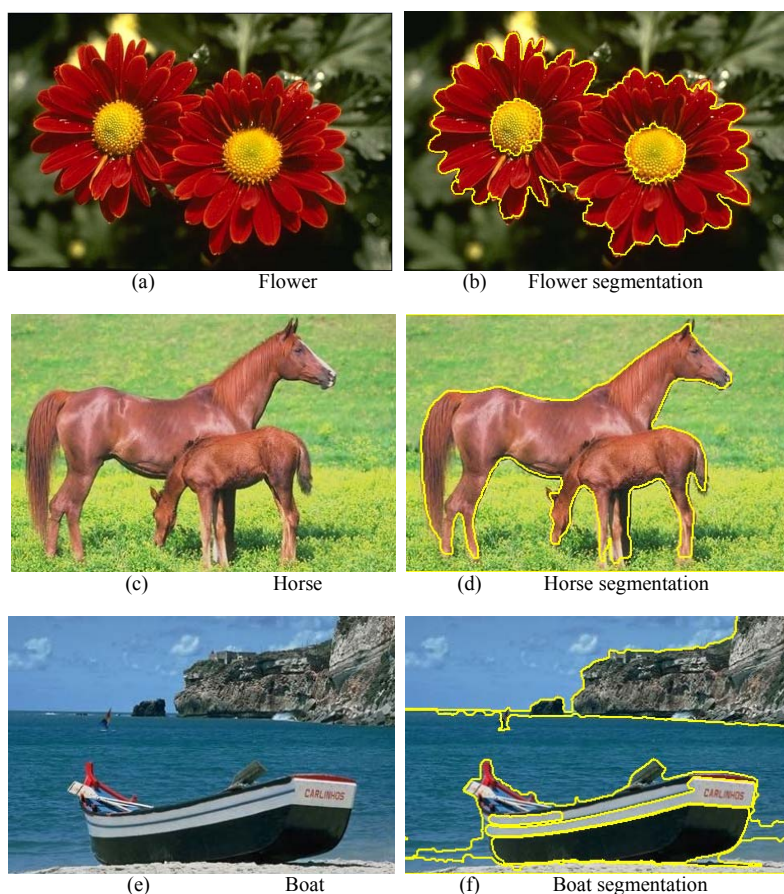


Figure 3. Color image segment results