

Text Segmentation from Images with Various Light Conditions Based on Gaussian Mixture Model

Khoa Anh Tran, Guesang Lee*

Department of Electronics and Computer Engineering
Chonnam National University, Gwangju, 500-757, Korea

ABSTRACT

Standard Gaussian Mixture Model (GMM) is a well-known method for image segmentation. However, one of its problems is that we consider the pixel as independent to each other, which can cause the segmentation results sensitive to noise. It explains why some of existing algorithms still cannot segment texts from the background clearly. Therefore, we present a new method in which we incorporate the spatial relationship between a pixel and its neighbors inside 3x3 windows to segment the text. Our approach works well with images containing texts, which has different sizes, shapes or colors in case of light changes or complex background. Experimental results demonstrate the robustness, accuracy and effectiveness of the proposed model in image segmentation compared to other methods.

Keywords: Gaussian Mixture Model (GMM), Image Segmentation, Spatial Neighboring Relationships, Expectation Maximization.

1. INTRODUCTION

TODAY, with the development of information technology, scanned images are widely used in many fields such as office document, digital library and electronic commerce. Many works have been done on text segmentation in the past few years. These works can be broadly classified into three major categories methods: color clustering which based on repetitive color using color information [8], [9]; segmentation based on histogram threshold [1], [10], and statistical models method [11], [12], [13], [14] which are very popular today.

The recent methods remain several problems. The color clustering techniques assume that there are several colors in images and pixels with the similar colors, which can be classified into the same category. Furthermore, the performance is only good when the color of text pixels is uniform. Another drawback is that it is too sensitive to noises and text resolution. The threshold method also has its problems. It cannot separate text pixels clearly, when background and text are in similar colors. Statistical model method builds a model for text pixels or all the pixels in images, and then classifies each pixel into different categories.

For previous algorithms, there are some problems need to be solved:

- 1) Inner area of small objects such as symbols is often too small to form basins used in the segmentation.
- 2) It is impossible to detect boundaries of objects with ramp,

roof and step solve these problems and to develop an efficient segmentation method for color images from journals and newspapers as well as others images of this kind.

3) Problems due to lighting conditions, variation in front shape and color, shadow of images.

Recently, GMM has become the famous algorithm for segment objects. However, one of the problems of standard GMM is that the spatial relationships between the neighboring pixels are not taken into its account. Although the GMM is a well-known and simple method for image segmentation, its segmentation result is thus sensitive to noise.

In this paper, we focus on natural scene texts and develop a color segmentation method for OCR systems. As we know, most of typical images in natural scenes, scanned pages of color journals and newspapers contain different types of texts. In our case, we can consider texts as objects, which have clearly defined edges and should be segmented from the background. We present a solution to handle the problem of other methods based on Gaussian Mixture Model, which currently remains as a popular method for text image segmentation [15].

In this method, the first step we calculated the local spatial relationship between neighboring pixels, the second step we modify Gaussian Mixture Model method and using expectation- maximization method estimate the parameters of GMM. The results is binarized images in which texts is segmented completely from complex background.

The remainder of the paper is organized as follows. In Section 2 we introduce again standard GMM. Section 3 we describe the details of proposed method. Section 4 is our experimental results based on the challenging ICDAR 2003

* Corresponding author, Email : gilee@chonnam.ac.kr
Manuscript received Nov. 16, 2012; revised Mar 10, 2013;
accepted Mar 20, 2013

word dataset and finally conclude the work in Section V. Finally, we make conclusion in section 5.

2. STANDARD GAUSSIAN MIXTURE MODELS

As we know, the standard GMM considers each pixel in image as observation sample x_i and the density function at an observation is:

$$p(x_i) = \sum_{k=1}^K \pi_k p(x_i | \Omega_k) \quad (1)$$

Where π_k is the prior distribution indicates how likely pixel x_i belongs to the class Ω_k , and π_k must satisfy the constraints $0 \leq \pi_k \leq 1$ and $\sum_{k=1}^K \pi_k = 1$, here K is the number of Gaussian Models. In our case, we assume that all texts in images have the same color and background usually has color that is different from the texts. Therefore, we choose K= 2 in order to making mask image which indicates which part is either foreground or background separately.

The Gaussian distribution $p(x_i | \Omega_k)$ is given by

$$p(x_i | \Omega_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}\right) \quad (2)$$

Where μ_k and σ_k is the mean and covariance of the Gaussian distribution $p(x_i | \Omega_k)$.

Our aim is find parameters $\Theta = (\mu_k, \sigma_k, \pi_k)$ in order to maximize the likelihood function:

$$L(\Theta) = \sum_{i=1}^N \log\left(\sum_{k=1}^K \pi_k p(x_i | \Omega_k)\right) \quad (3)$$

There are many techniques have been developed to determine these parameters. One of the popular method is expectation maximization algorithm which based on Bayes' theorem. We will illustrate this algorithm more details for our purpose in Section 4.

It is well known that the neighbor pixels within a small window are usually similar. One of the major drawbacks of standard GMM is that the pixel observations x_i are consider to be independent to each other so that the information of the neighboring pixels is not taken into account. Another limitation of standard GMM is that the prior distribution π_k has for each class the same weigh values for each pixel in the image, and thus, it sometimes misclassifies the observation into the correct class.

3. THE PROPOSED METHOD

3.1 Overview of system

In this section, we introduce the method to handle some

previous problems of standard GMM. The method has been proposed in [17]. However, they also use the gradient method instead of using expectation-maximization algorithm to maximize the log-likelihood function. In our case, we simply using that modified GMM, combined with traditional expectation maximization algorithm to segment texts image in natural scenes. These images are influenced by lighting or noises, which can cause the complex textures, and thus, the standard GMM fails to detect texts. The main difference of this method from the standard GMM is that the prior distributions π_{ik} are different for each pixel, which is indicated by the values of itself and the neighbor pixels. The overview of the proposed system is illustrated in Fig. 1.

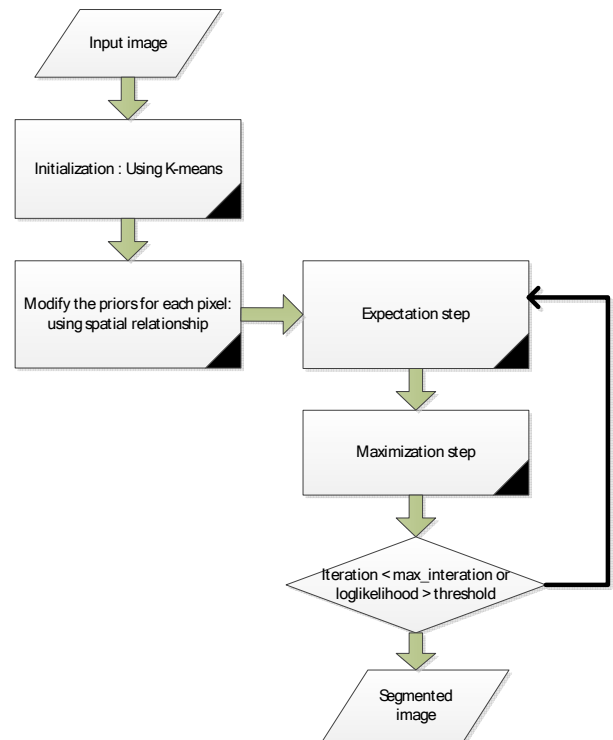


Fig. 1. Flowchart of our system.

3.2 The local spatial relationship

In the standard GMM method, the prior for each pixel in initialization step of expectation-maximization algorithm is equal because pixels are considered as there are no relations between them. For example, with K = 3, the prior for each pixel is (1/3, 1/3, 1/3) which means that the percentage of pixel belongs to each class is the same. Therefore, we easily misclassify pixels in case of complex textures. Some noise pixels in the background, which have same color as texts also classify as texts de furthermore, there is high time consuming for convergence.

By using spatial relationship, we could determine the prior for each pixel in initialization step. Each pixel should have the prior for each class differently. For example, pixels in texts should have high prior for class foreground and lower priors for others while pixels in background, which even have the same color with texts because of noises must have low prior for

foreground class and higher prior for other classes. This is the main ideal of the proposed method.

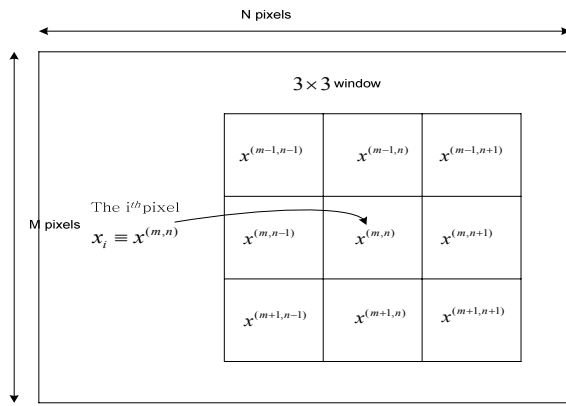


Fig. 2. 3x3 window including center pixel and its neighborhood.

The prior distribution of each pixel are determined in the following steps:

Step 1: In the images of size \$M \times N\$, we replace the value of each pixel with the mean of its neighborhoods inside 3x3 window.

$$\bar{x} = \overline{x^{(m,n)}} = \frac{1}{9} \sum_{u=-1}^1 \sum_{v=-1}^1 x^{(m+u,n+v)} \quad (4)$$

Step 2: Define \$w_k(\bar{x}_i)\$ to represent the weight of each pixel in each class \$\Omega_j\$ as :

$$w_k(\bar{x}_i) = w_k(\overline{x^{(m,n)}}) = \exp\left(-\frac{\left(\overline{x^{(m,n)}} - \mu_k\right)^2}{2\sigma_k^2}\right) \quad (5)$$

Where \$\mu_k\$ and \$\sigma_k\$ are the means and covariance for classes that are the output of K-means clustering algorithm. We recognize that the neighboring pixels have strong contribution to determine the prior for the center pixel. In case of pixels belong to the text, which usually have similar colors, they will have high weight for foreground class while many separated noise pixels will have lower weight for this class.

Step 3: In this step, we already have the weight matrix of size \$M \times N\$. In order to differentiate the pixels inside the text and the ones in background, we take average of the weight matrix inside of 3x3 windows again.

$$\overline{w_k(\bar{x}_i)} = \overline{w_k(\overline{x^{(m,n)}})} =$$

$$\frac{1}{9} \sum_{u=-1}^1 \sum_{v=-1}^1 \exp\left(-\frac{\left(\overline{x^{(m+u,n+v)}} - \mu_k\right)^2}{2\sigma_k^2}\right) \quad (6)$$

Finally, we can determine the prior distribution for each pixels corresponding to each class \$\Omega_j\$.

$$\pi_{ik} = \frac{\overline{w_k(\bar{x}_i)}}{\sum_{j=1}^K \overline{w_j(\bar{x}_i)}} \quad (7)$$

The idea to incorporate the spatial constraints in our method is based on a fact that neighboring pixels in an image are similar in some sense. Based on this relationship, we replace each pixel value in an image with the average value of its neighbors, including itself.

3.3 Expectation maximization algorithm

Expectation maximization algorithm was taken to estimate the parameters of Gaussian mixture models. If all samples are independent and identically distribution, then the log-likelihood functions is defined as standard GMM.

There are two steps:

Expectation step: Compute and uniform the posterior probability of each sample \$x_n (1 \le n \le N)\$ belonging to k-th cluster as follow:

$$p(\Omega_k | x_i) = \frac{\pi_k p(x_i | \Omega_k)}{\sum_{k=1}^K \pi_k p(x_i | \Omega_k)} \quad (8)$$

Where \$p(x_i | \Omega_k)\$ is computed with Eq. (2).

Maximization step: Get the new parameters throughout the maximum of Eq. (3). The concrete equations as follow:

$$\pi_k = \frac{1}{N} \sum_{n=1}^N p(\Omega_k | x_n) \quad (9)$$

$$\mu_k = \frac{\sum_{i=1}^N p(\Omega_k | x_i) x_i}{\sum_{i=1}^N p(\Omega_k | x_i)} \quad (10)$$

$$(\sigma_k)^2 = \frac{\sum_{i=1}^N p(\Omega_k | x_i) (\|x_i - \mu_k\|)^2}{\sum_{i=1}^N p(\Omega_k | x_i)} \quad (11)$$

When EM algorithm is convergent, the posterior probability of each sample \$x_i (1 \le i \le NxM)\$ belonging to each \$k (1 \le k \le K)\$ cluster is taken. Then the class of each sample can be determined means of using the maximum probability rule.

4. EXPERIMENTAL RESULTS

4.1 Data set

In this part, we use sample images from the ICDAR 2003 Robust Word Recognition dataset [16] for our experiments. It consists of 171 natural scene text images. The data set contains different sizes and labels. These images have several degradations due to uneven lighting, complex background, blur and similar foreground background colors. Fig. 2 shows some examples from the image data set.



Fig. 3 . Some samples images we considered in this work.

4.2 Experiments and result

Two evaluation methods are used to evaluate text segmentation performance. One is pixel classification rate (PCR), which is computed based on pixels and widely used in image segmentation tasks. The other is character recognition (CRR), we calculate based on characters and popularly used in text segmentation. They are defined such as:

$$PCR = \frac{\text{correctly classified pixels}}{\text{total pixels}}$$

$$CRR = \frac{\text{correctly recognized characters}}{\text{total characters}}$$

To evaluate the performance of the proposed method based on the extended GMM method, we compare it with the well-known methods based on threshold histogram: Niblack [1], Kittler [11], K-means and standard Gaussian Mixture color clustering method in chapter book [2] and produce successful segmentation results. The PCR and CRR of four methods are shown in Table 1. The proposed method performs the best among the four methods. Some segmentation examples are also illustrated in Fig. 4, which shows that compared with other four methods the proposed method can obtain better segmentation results from images with complex background.

Table 1. Segmentation results with respect to well-known valuation measures (average).

Method	CRR	PCR
Niblack	0.77	0.85
Kittler	0.74	0.86
K-means	0.76	0.87
GMM	0.75	0.85
Proposed	0.89	0.91



Fig. 4. (a) Original image, (b) Segmentation result of our method.

5. CONCLUSION

In this article, we present a method to incorporate the spatial relationship between neighboring pixels into the standard GMM. This method can segment ramp, roof edges, and different size text in complex background from scanned image efficiently. This method solved some problems such as inner area of small text, variation in the front shape, color and shadow of images based on extended Gaussian mixture model with EM algorithm. The proposed method can handle well with scanned images from journals and newspapers and compare the proposed approach with other four text segmentation methods. Experiment results demonstrate that our method gives better text segmentation results and can effectively segment text from images with complex background. In the future, we should investigate more about GMM clustering in text segmentation.

ACKNOWLEDGEMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technologies(2012-047759) and the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2012-H0301-12-3005).

REFERENCES

- [1] W.Niblack, *Introduction to Digital Image Processing*, Prentice Hall, 1986, pp.115-116.
- [2] Omer. D, A. H. Musa, Sahoo. K. Prasanna, *Image processing with Matlab, Applications in Medicine and Biology*, CRC Press, 2009, pp 269-284.
- [3] Carlo Tomasi, *Estimating Gaussian Mixture Densities with EM – A Tutorial*, Duke University.
- [4] C. M. Bishop, *Pattern Recognition and Machine Learning*, Berlin, Germany: Springer, 2006.
- [5] G. J. MacLachlan and T. Krishnan, *The EM Algorithm and Extensions*, Wiley Series in Probability and Statistics. New York: Wiley, 1997.
- [6] Gonzalez, Rafael; Richard Woods. *Digital Image Processing (3rd ed.)*. Upper Saddle River, New Jersey: Pearson Education, Inc.. pp. 165–68. ISBN 978-0-13-168728-8.

Original	Nblack	Kittler	K-means	GMM	Our Method

Fig. 5. Comparison of the proposed method and previous methods.

- [7] D. Marr and E.Hildreth. *Theory of edge detection*. Proc. R. Soc. Lond. A, Math. Phys. Sci, vol. B 207, 1980, pp. 187-217.
- [8] H. D. Cheng, X. H. Jiang, Y. Sun, Jingli Wang, "Color image segmentation; Advances and prospets", *Pattern Recognition* 34.
- [9] F. Meyer, "Color image segmentation", *Proc. IEE Int. Conf. Image Processing and its Applications*, The Netherlands, 1992, pp. 303-306.
- [10] N. Otsu, "A threshold selection method from gray-level histograms", *IEEE Transactions Systems Man Cybernetics*, vol. 9, no. 1, 1979, pp: 62-66.
- [11] J. Kittler, J. Illingworth, and J. Foglein, "Threshold selection based on a simple image statistic", *Computer Vision, Graphics, and Image Processing*, vol. 30, no. 2, 1985, .
- [12] D. Ziou, S. Tabbone, "Edge Detection Techniques-An Overview", technical report, No. 195, Dept Math & Informatique, Universit de Sherbrooke, 1997.
- [13] Stephen G, Tianshi G, Daphne K, "Region-based Segmentation and Object Detection", *NIPS conference*, 2009.
- [14] Chan, K.C.C, Xudong H, Bao. P, "Fuzzy segmentation for document analysis", *IEE Int. Conf. System, Man, and Cybernetics, Computational Cybernetics and Simulation*, vol. 2, 1997, pp. 977-982.
- [15] Gudnason, J., Brookes. M, "Distribution based classification using Gaussian Mixture Models", *IEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, 2002, pp. IV-4159.
- [16] B. Gatos, K. Ntirogiannis, and I. Pratikakis, "ICDAR 2009 document image binarization contest (DIBCO 2009)," *ICDAR*, 2009, pp. 1375-1382.
- [17] Thanh Minh Nguyen, Wu, Q.M.J, " Gaussian-Mixture-Model-Based Spatial Neighborhood Relationships for Pixel Labeling Proble", *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2012, pp. 193-202.



Tran Anh Khoa

He received the B.S and M.S degree in Faculty of Network and Communication Systems from Moscow Technical University of Communication and Informatics, Russia. He is currently a Ph.D student in Department of Electronics and Computer Engineering,

Chonnam National University, Korea. His research interests are mainly in the field of image processing, computer vision.



Guesang Lee

He received the B.S. degree in Electrical Engineering and the M.S. degree in Computer Engineering from Seoul National University, Korea in 1980 and 1982, respectively. He received the Ph.D. degree in Computer Science from Pennsylvania State University in 1991.

He is currently a professor of the Department of Electronics and Computer Engineering in Chonnam National University, Korea. His research interests are mainly in the field of image processing, computer vision and video technology.