Image Segmentation Using Expectation Maximization and Its Application to Digital Copying

Mohamed N. Ahmed, Software Research, Lexmark International, Inc., Lexington, Kentucky, USA

Abstract

In this paper, we present a new system to segment and label document images into text, halftone images, and background using feature extraction and unsupervised clustering. Each pixel is assigned a feature pattern. The invariant feature pattern is then assigned to a specific region using the Expectation-Maximization (EM) algorithm. Once the segmentation is performed, a specific enhancement filter can be applied to each document component.

Introduction

Digital copying, in which a digital image is obtained from a scanning device and then printed, involves a variety of inherent factors that compromise image quality. Ordered halftone patterns in the original document interact with the periodic sampling of the scanner, producing objectionable moiré patterns. These are exacerbated when the copy is reprinted with an ordered halftone pattern. In addition, limited scan resolution blurs edges, degrading the appearance of detail such as text. Fine detail also suffers from flare, caused by the reflection and scattering of light from the scanner's illumination source. Flare blends together nearby colors, blurring the high-frequency content of the document.

To suppress moiré, a filter may be constructed that is customized to the frequencies of interest. However, both the detection of the input halftone frequencies and the frequency-domain filtering itself can require significant computational effort. Although crude, a simple, small low-pass filter can correct the majority of moiré artifacts. Unfortunately, low-pass filtering affects detail as well, blurring it even further.

Sharpening improves the appearance of text and fine detail, countering the effects of limited scan resolution and flare. Edges become clear and distinct. Of course, other artifacts such as noise and moiré become sharper as well.

The solution is simple in concept: determine the content of regions within the scanned image and then apply the appropriate filter to each region. Sharpening should be performed on fine detail, while moiré suppression should be applied to certain periodic artifacts.

From the above discussion, we can conclude that for an image enhancement system to work properly, a preprocessing step should include a segmentation of the document into text, halftone and background. If this step is successfully completed, the application of an appropriate filter should be straightforward.

Several approaches for document segmentation have been proposed.[3-8] These techniques can be broadly classified as bottom-up or top-down. Bottom-up methods start from the pixel level and merge regions together into larger and larger components. Top-down techniques apply a priori knowledge about the page to hypothesize and split the page into blocks which are subsequently identified and further subdivided. Top-down approaches work well with pre-specified layouts such as technical papers. However, the performance of these techniques degrades significantly when different components are touching or overlapping. Among bottom-up approaches, texture-based schemes have attracted much attention.[6-8]

These methods treat different components of a document image as different textures. The scanned document images are convolved with a set of masks to generate feature vectors. Each feature vector is then classified into different classes using a pre-trained classifier. One problem associated with these approaches is the mask size for extracting local features. If the mask size is too small, it is difficult to detect large scale textures such as large fonts. On the contrary, if a large mask is chosen, the computational complexity will increase dramatically.[8]

In this paper, we propose a simple document segmentation technique that involves extracting discriminating features and clustering them into different regions using the Expectation-Maximization (EM) algorithm.

System Description

The presented method consists of the following steps: feature extraction and clustering and labeling.

Feature Extraction

The first step in our approach is to extract a feature pattern for each pixel in a gray level image. A feature vector, $\mathbf{y}$, is a set of measurements $\{y^1, y^2, \cdots, y^d\}$ which condenses the description of relevant properties of the image into a small, Euclidean feature space of d dimension. The number of needed features depends on the complexity of the image. The components of the feature vector may include gray values, gray values through different filters, texture measures, Markov random field features, fractal dimension measures, and gradient magnitudes and directions.[7]

In this paper, we will construct our feature vector from measurements obtained from gray level distribution in a window, W, of dimension $w \times w$ centered around each pixel. These features describe the first order gray level distribution without considering the spatial interdependence. Two features were selected:

1. The mean gray level, μ,

$$\mu = \frac{1}{w^2} \sum_{(x,y) \in W} I(x, y) \tag{1}$$

where $I(x,y)$ is the gray level at location (x,y).

2. The variance σ^2 of the gray level variation

$$\sigma^2 = \frac{1}{w^2 - 1} \sum_{(x,y) \in W} (I(x, y) - \mu)^2. \tag{2}$$

Image Segmentation

In order to segment the image, we model the joint feature distribution with a mixture of Gaussians. We use the Expectation-Maximization (EM) algorithm to estimate the parameters of this model. The EM algorithm is used for finding maximum likelihood parameter estimates when there is missing or incomplete data. In our case, the missing data is the gaussian cluster to which the points in the feature space belong.[9]

Assuming that we use C clusters in the mixture model, then the joint distribution can be modeled as

$$p(y / \Theta) = \sum_{i=1}^{C} p(y, \theta_i) = \sum_{i=1}^{C} \pi_i p(y \mid \theta_i), \tag{3}$$

where y is the feature vector, π_i represents the weight of the i-th mixture, and

$$p(y \mid \theta_i) = \frac{1}{(2\pi)^{d/2} \det \sum_i^{1/2}} \exp\left[-\frac{1}{2} (y - \mu_i)^T \sum_i^{-1} (y - \mu_i) \right], \tag{4}$$

where μ_i and Σ_i are the mean and the covariance matrix for the i-th class.

Parameter Estimation

The EM algorithm to cluster N feature vectors iterates as follows:

1. The E-step:

For every pixel at location t, $1 \le t \le N$, compute δ_{it} as

$$\delta_{it}^k = \frac{\pi_i^k p(y_t \mid \Theta_i^k, C_i)}{\sum_{l=1}^{m} \pi_l^k p(y_t \mid \Theta_i^k, C_l)} \tag{5}$$

where y_t is the feature vector at location t, π_i^k is the mixing proportion of the i-th mixture at step k, and Θ_i^k is estimated parameter for the i-th mixture at step k.

2. The M-step:

We compute the new mean, the new variance and the new proportion from the following equation:

$$\pi_i^{k+1} = \frac{1}{N} \sum_{t=1}^{N} \delta_{it} \tag{6}$$

$$\mu_i^{k+1} = \frac{\sum_{t=1}^{N} \delta_{ti}^k y_t}{\sum_{l=1}^{C} \delta_{tl}^k} \tag{7}$$

$$\sum_i^{k+1} = \frac{\sum_{t=1}^{N} \delta_{tl}^k (y_t - \mu_i^k)(y_t - \mu_i^k)^T}{\sum_{l=1}^{C} \delta_{tl}^k} \tag{8}$$

3 Repeat steps 1 and 2 until the relative difference of the subsequent values of Eq. 6, Eq. 7, and Eq. 8 are sufficiently small.

Results

We tested our algorithm with several images scanned at 600 dpi. Each pixel was assigned a feature vector using a 5 × 5 window. As shown in Figures 1 and 2, the feature modeling step was performed using a mixture of 4 Gaussians. The parameters were estimated using the EM algorithm as described above. Figures 1 and 2 show the original histogram of the features, their mixing components and the final modeling. Once the modeling step is performed, document pixels are assigned to one of three classes: text, images, and background. Figures 3 and 4 illustrate such operation on two different scans. Figure 4 was further postprocessed to remove segmentation outliers.

Conclusions

In this paper, we presented a new technique to segment document images. With this technique, each pixel is assigned a feature pattern. The invariant feature pattern is then assigned to a specific region using the Expectation-Maximization algorithm. Once the segmentation step is performed, specific filters and interpolation functions can be applied to each document component.

Future work will seek the incorporation of an image modeling technique such as Markov random Field (MRF) to model spatial interactions between pixels. Such modeling should produce a segmentation that is more robust to noise.

References

1. M. Ahmed, B. E. Cooper, and S. Love, “Document Image Segmentation Using a Two-stage Neural Network,” *Proc. of SPIE* , Vol. 3962, pp. 25-33, January 2000.
2. M. Unser, A. Aldroubi, and M. Eden, “Enlargement or Reduction of Digital Images with Minimum Loss of Information,” *IEEE Trans. on Image Processing* , Vol. 4, No. 3, pp. 247-258, March 1995.
3. Y. Y. Tang, C. Y. Suen, and M. Cheriet, “Document Analysis and Understanding: A Brief Survey,” *Proc. Int. Conf. on Document Analysis and Recognition*, pp. 17-31, 1991.

4. H. Cheng and C. Bouman, "Trainable Context Model for Multiscale Segmentation," *Proc. Int. Conf. on Image Processing (ICIP)*, October 1998.
5. L. O'Gorman, "The Document Spectrum for Page Layout Analysis," *IEEE Trans. on Pattern Matching and Machine Intelligence (PAMI)*, Vol. 15, No. 11, November 1993.
6. K. Etemad, D. Doermann, and R. Chellapa, "Page segmentation using decision integration and wavelet packets," *Proc. Int. Conf. of Pattern Recognition*, Vol. 2, pp. 345349, 1994.
7. A. K. Jain and Y. Zhong, "Page segmentation using texture analysis," *Pattern Recognition*, Vol. 29, No. 5, pp. 343770, 1996.
8. H. Cheng, C. Bouman, and J. Allebach, "Multiscale Document Segmentation," *Proc. of IS&T 50th Annual Conference*, pp. 417-425, May 1997.
9. C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation Using Expectation-Maximization and its Application to Image Query," *IEEE Trans. on Pattern Matching and Machine Intelligence (PAMI)* , Vol. 24, No. 8, pp. 1026-1038, August 2002.
10. A. P. Witkin, "Scale space filtering," *Proc. International Joint Conference on Artificial Intelligence*, pp. 1019-1023, 1983.
11. S. Haring, M. Viergever, and J. Kok "A multiscale approach to image segmentation using Kohonen networks," *Proc. IPMI*, pp. 212-224, Berlin, 1993.
12. J. J. Koenderink, "The structure of images," *Biological Cybernetics*, Vol. 50, pp. 363-370, 1984.
13. L. Florack and B. M. Romeny, "Scale and the differential structure of images," *Image and Vision Computing*, Vol. 10, pp. 376-388, 1992.
14. R. M. Haralik and L. G. Shapiro, "Computer and Robot Vision," Addison-Wesley Publishing Company, 1992.
15. S. Dellepiane et al., "Model generation and model matching of real images by fuzzy approach," *Pattern Recognition*, Vol. 25, No. 2, pp. 115-137, 1992.
16. S. Haykin, "Neural Networks: A Comprehensive Foundation," Macmillan College Publishing Company, 1994.
17. T. Kohonen, "Self Organization and Associative Memory," Springer-Verlag, 1984.

Author Biography

Dr Mohamed Nooman Ahmed received a B.Sc. and M.Sc. degree in Biomedical Engineering from Cairo University, Egypt in 1989 and 1992 respectively, and a Ph.D. in Computer Engineering from the University of Louisville, Kentucky, in 1998. Since 1999, he has been an Imaging Research Scientist at Lexmark International Inc. Dr. Ahmed's main research interests include image processing, segmentation, visualization, Fuzzy logic, and Neural networks.

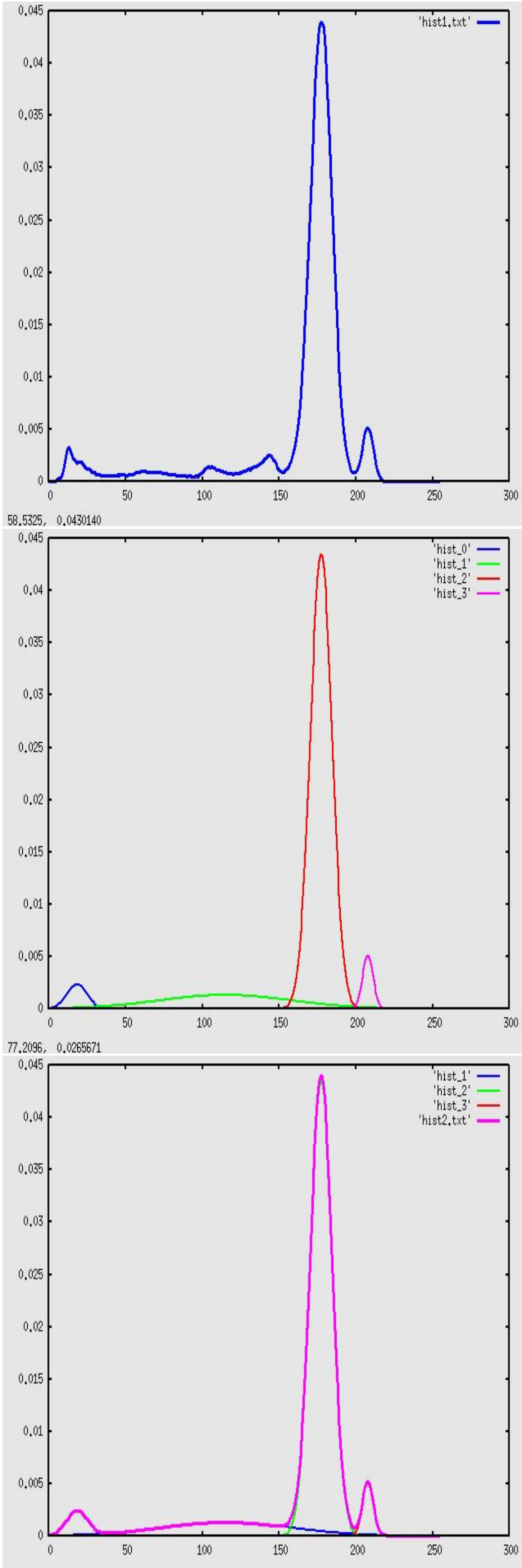

Figure 1: *Histogram of the mean feature and its modeling using 4 gaussian mixtures. Middle figure shows the different components and Bottom gur e shows their combination to model the histogram.*

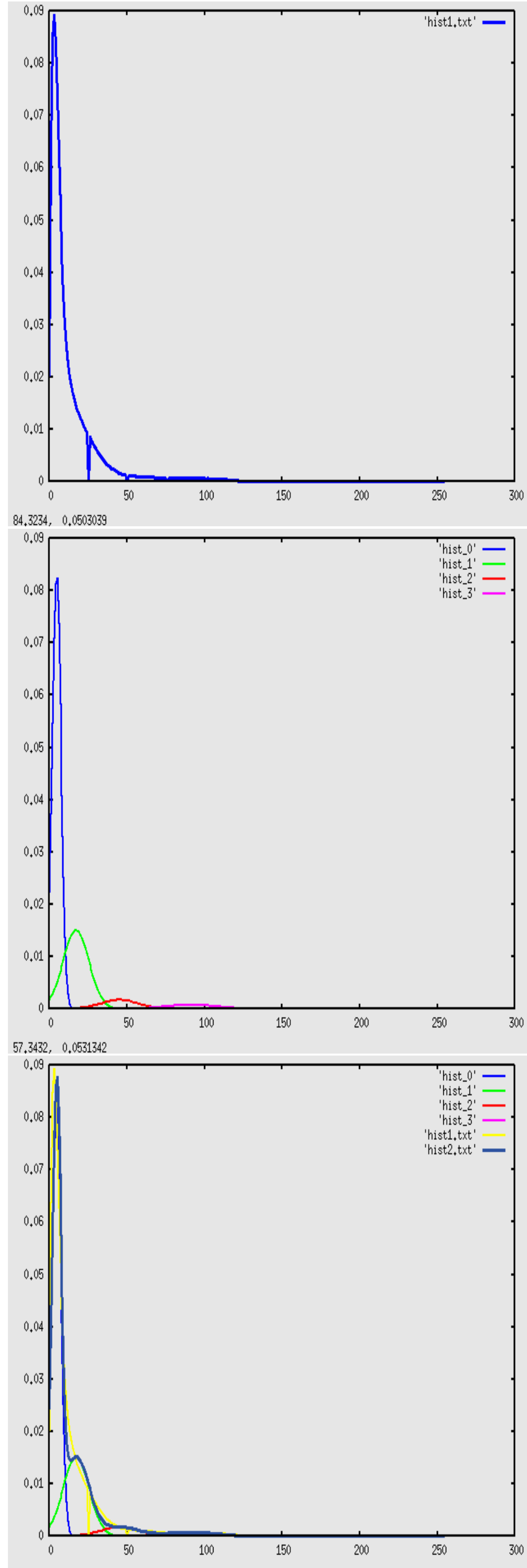

Figure 2: *Histogram of the variance feature and its modeling using 4 gaussian mixtures. Middle figure shows the different components and Bottom gur e shows their combination to model the histogram.*

Figure 3: Original and segmented document into text (back), image (gray), and background (gray)

Figure 4: Original and segmented document into text (black), image (gray), and background (gray)