Projection-based Scanned Image Enhancement

Mohamed Nooman Ahmed,Ph.D.
Software Research, Lexmark International, Inc.
Lexington, KY, USA

Abstract

In this paper, we present a new system to segment and label document images by combining statistical and multiscale view of different image components. Texture of text, halftone and images are characterized by modeling the distribution of a novel intensity projection technique using a mixture of K Gaussians. Model parameters are then estimated using the expectation maximization (EM) algorithm. Using the proposed algorithm, halftone areas were successfully differentiated from text regions

Introduction

Document images typically contain a combination of text, background, and halftoned images. Text, including line art and similar graphical content, is characterized by sharp, high-contrast edges and thin strokes. The background of the document is usually white or nearly white, and it normally has a smooth texture. Halftoned images consist of a pattern of small dots. In most cases, the dots are arranged in an ordered pattern and will vary slightly in size according to the darkness of the image they represent.

When documents are scanned and then printed, the resultant digitally copied documents usually suffer from blurring, flare, noise, and moire [1]. To address these distortions, several techniques have been presented to apply content-based filtering. In these methods, a segmentation step is performed to classify each pixel into text, background, and halftoned image regions. Appropriate filters are then applied to each document component [1]-[7].

Discriminating between text and halftone dots, especially low frequency halftones, has proved to be problematic [1]. Edge of halftone dots are often miss-classified as text which increases moire. The simplest way of suppressing halftone noise is to smooth the entire scanned document with a single low pass filter, a process known as descreening. However, this approach also softens the image edges.

To resolve this problem, Queiroz et al [3] presented an approach for descreening based on wavelet decomposition. In this approach, wavelet decomposition of the halftone image facilitates a series of spatial and frequency selective processing to preserve most of the original image contents while eliminating the halftone noise. Siddiqui et al. [4] recently presented a training based descreening approach that attempts to descreen the image while preserving edges by combining two non-linear image processing techniques, Resolution Synthesis based Denoising (RSD) and Modified SUSAN filtering. Unfortunately, the application of these techniques does not preserve fine details.

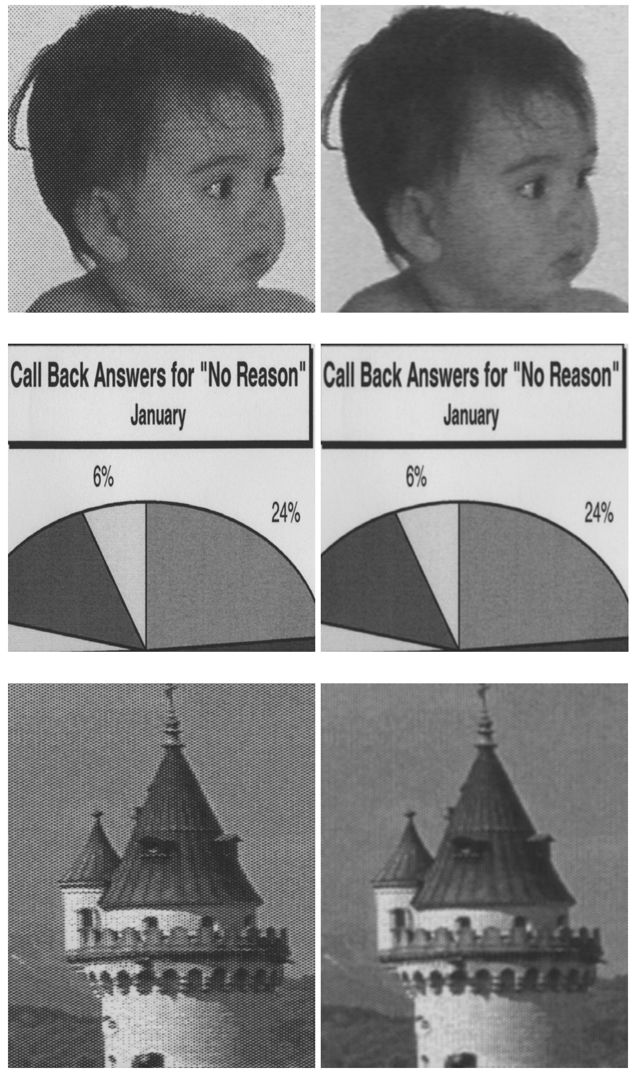

Figure 1. *Examples of portions of scanned documents containing 106 lpi screens halftones and the blurring effect that results from applying a simple lowpass filter to smooth the entire document*

In this paper, we present a new system to segment and label document images and robustly discriminate between halftone areas and text by combining statistical and multiscale view of different image components. Texture of text, halftone and images are characterized by modeling the distribution of a novel intensity projection technique using a mixture of K Gaussians. Model parameters are then estimated using the expectation maximization (EM) algorithm. Using the proposed algorithm, halftone areas were successfully differentiated from text regions.

Intensity Projection

To measure the degree of uniformity, we will first extract horizontal, vertical, and diagonal projections from an $W \times W$ window centered around each pixel

$$H_i = \sum_{y=0}^{W-1} f(i,y) \qquad i = 0,1,\cdots,W-1 \tag{1}$$

$$V_i = \sum_{x=0}^{W-1} f(x,i) \qquad i = 0,1,\cdots,W-1 \tag{2}$$

$$D_1 = \sum_{y=0}^{W-1} f(i,i) \tag{3}$$

$$D_2 = \sum_{y=0}^{W-1} f(i,N-1-i) \tag{4}$$

We then compute the differences between successive vertical and horizontal projections and the difference between diagonal projections as follows

$$P_{hi} = H_{i+1} - H_i \qquad i = 0,1,\cdots,W-2 \tag{5}$$

$$P_{vi} = V_{i+1} - V_i \qquad i = 0,1,\cdots,W-2 \tag{6}$$

$$P_d = D_1 - D_2 \tag{7}$$

P_{hi}, P_{vi}, and P_d are combined into a feature vector $\mathbf{x}$. N samples of such vectors are obtained from a collection of scanned data that contains text and non-text (including halftone) areas. The distribution of this mixture of sample vectors is modeled as a mixture of Gaussians and its parameters are estimated using the Expectation-Maximization algorithm as described in the next sections.

Mixture Model

Assuming that we use K clusters in the mixture model, then the form of the probability density function is as follows:

$$p(\mathbf{x}|\Theta) = \sum_{i=1}^{K} p(\mathbf{x},\theta_i) = \sum_{i=1}^{K} \pi_i p(\mathbf{x}|\theta_i), \tag{8}$$

where $\mathbf{x}$ is the feature vector and π_i represents the weight of the i-th mixture. The symbol Θ represent the parameter set $\{\pi_1, \pi_2, \cdots, \pi_K, \theta_1, \theta_2, \cdots, \theta_K\}$, and $p(\cdot)$ is a d-variate Gaussian density parametrized by θ_i

$$p(\mathbf{x}|\theta_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu_i)^T \Sigma_i^{-1}(\mathbf{x}-\mu_i)}, \tag{9}$$

where μ_i and Σ_i are the mean and the covariance matrix for the i-th class [9].

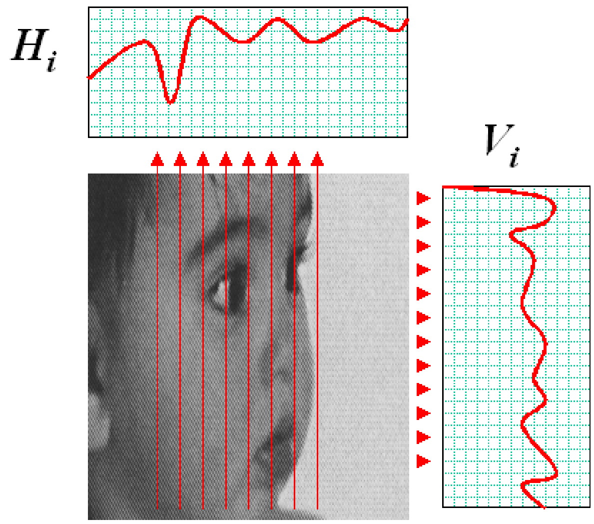

Figure 2. *Horizontal and vertical intensity projections of an input window.*

Given a set of N independent and identically distributed samples $\mathscr{X} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(N)}\}$, the log-likelihood corresponding to a mixture is

$$log\ p(\mathscr{X}|\Theta) = log \prod_{n=1}^{N} p(\mathbf{x}^{(n)}|\Theta) \tag{10}$$

The Maximum-Likelihood estimate (MLE) $\hat{\Theta}$ is given by

$$\hat{\Theta} = arg \max_{\Theta}\{log\ p(\mathscr{X}|\Theta)\} \tag{11}$$

In this paper, $\hat{\Theta}$ is estimated iteratively using the Expectation-Maximization algorithm, as outlined in the following section.

Expectation-Maximization Algorithm

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

- **The E-step**: For every pixel at location t, $1 \le n \le N$, compute δ_{in}^t as

  $$\delta_{in}^t = \frac{\pi_i^t p(\mathbf{x}^{(n)}|\Theta_i^t, K_i)}{\sum_{j=1}^{C} \pi_j^t p(\mathbf{x}^{(n)}|\Theta_j^t, K_j)} \tag{12}$$

  where $\mathbf{x}^{(n)}$ is the n-th feature vector, $\pi_i^t$is the mixing proportion of the i-th mixture at step t, and Θ_i^t is estimated parameter for the i-th mixture at step t.

- **The M-step**:we compute the new mean, the new variance and the new proportion from the following equation:

  $$\pi_i^{t+1} = \frac{1}{N}\sum_{n=1}^{N} \delta_{in}^t \tag{13}$$

  $$\mu_i^{t+1} = \frac{\sum_{n=1}^{N} \delta_{in}^t \mathbf{x}^{(n)}}{\sum_{j=1}^{K} \delta_{jn}^t} \tag{14}$$

$$\Sigma_i^{t+1} = \frac{\sum_{n=1}^{N} \delta_{in}^t (\mathbf{x}^{(n)} - \mu_i^t)(\mathbf{x}^{(n)} - \mu_i^t)^T}{\sum_{j=1}^{K} \delta_{jn}^t} \tag{15}$$

- Repeat steps 1 and 2 until the relative difference of the subsequent values of Eq. 13, Eq. 14, and Eq. 15 are sufficiently small.

Results

We tested our algorithm with 50 images scanned at 600 $\times$ 300 dpi. The images contained a mixture of text, images and halftones at different frequencies. We applied intensity projections at every pixel using an 11×11 window. The feature modeling step was performed using a mixture of 4 Gaussians. The parameters were estimated using the EM algorithm as described in section .

Figure 3 shows an original scan and its segmentation map. Text pixels are represented by red, image and background pixels are blue, while halftone pixels are yellow. Figures 4 and 5 present the results of applying the projection method to enhance different scans. The results demonstrate the ability of the technique in identifying text and fine details regions and in discriminating them from halftone pixels. As shown in Figures 4 and 5, text and fine details components can be sharpened effectively while halftone pixels can be descreened using a lowpass filter.

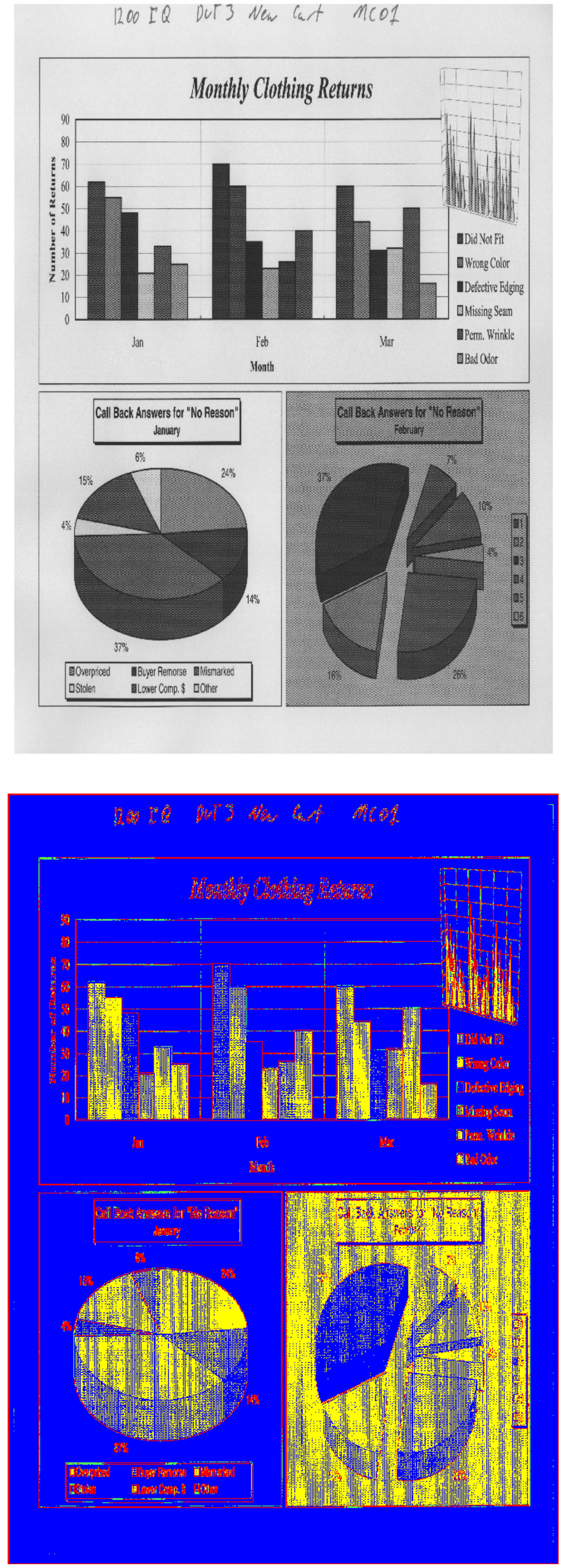

Figure 3. Original and segmented scanned document using the projection method. Text pixels are represented by red, image and background pixels are blue, while halftone pixels are yellow.

References

[1] M. N. Ahmed and Ahmed Eid, "Parameter Optimization for content-based image enhancement," *Proc.of the International Conference on Digital Printing Technologies (NIP 23)* , Alaska, September 2007.

[2] M. N. Ahmed and Brian E. Cooper, "Segmentation and Enhancement of Digital Copies Using a New Fuzzy Clustering Method," *Proc. of Electronic Imaging conference*, San Jose, CA, January 2006.

[3] J. Luo, R. Queiroz , and Z. Fan, "A Robust Technique for Image Descreening Based on the Wavelet Transform," *IEEE Trans. on Signal Processing* , Vol. 46, No. 4, 1998.

[4] S. Hasib and C. Bouman "Training-based descreening," *IEEE Trans. on Image Processing* , Vol. 16, No. 3, pp. 789-802, 2007.

[5] S. Kumar, R. Gupta, N. Khanna, S. Chaudhury, and S. Hoshi, "Text Extraction and Document Image Segmentation Using Matched Wavelets and MRF Model," *IEEE Trans. on Image Processing* , Vol. 16, No. 8, pp. 2117-2128, August 2007.

[6] 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.

[7] 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.

[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. Reddy, H. Chiang, and B. Rajaratnam, "TRUST-TECH-Based Expectation Maximization for Learning Finite Mixture Models," *IEEE Trans. on Pattern Matching and Machine Intelligence (PAMI)* , Vol. 30, No. 7, pp. 1146-1157, July 2008.

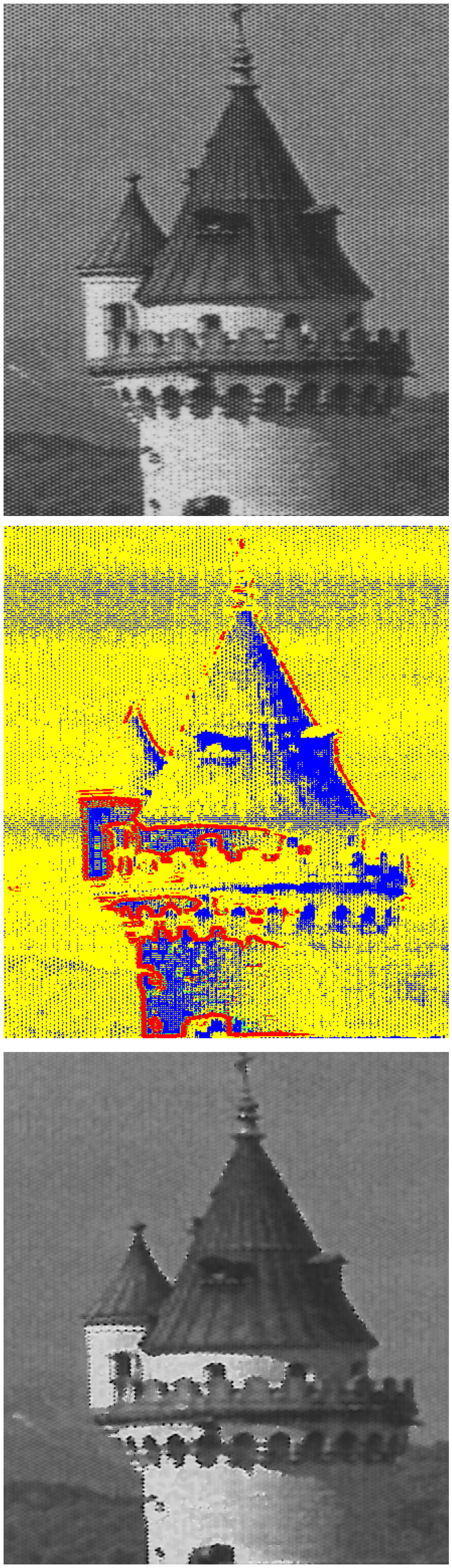

Figure 4. *Results from applying the projection-based method: (Top) Original test image, (middle) classification map, and (bottom) the enhanced image*

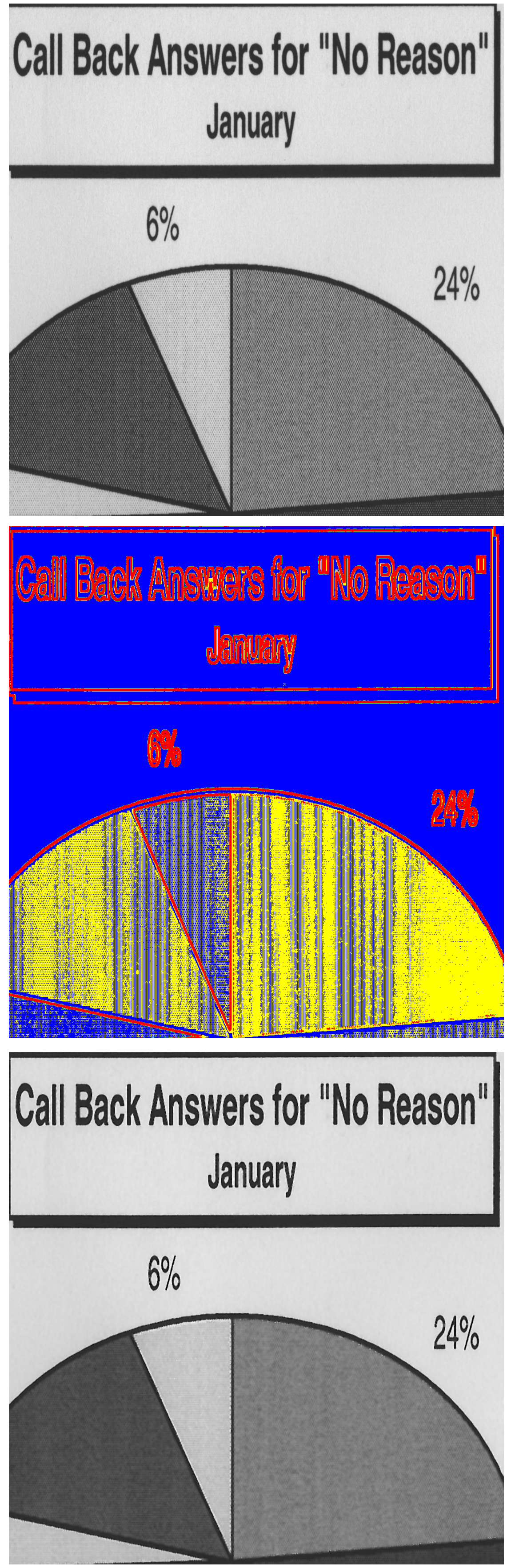

Figure 5. *Results from applying the projection-based method: (Top) Original test image, (middle) classification map, and (bottom) the enhanced image*