Degraded Document Image Enhancement

G. Agam[a], G. Bal[a], G. Frieder[b], O. Frieder[a]

[a]Illinois Institute of Technology, Chicago, IL 60616
[b]The George Washington University, Washington, DC 20052

ABSTRACT

Poor quality documents are obtained in various situations such as historical document collections, legal archives, security investigations, and documents found in clandestine locations. Such documents are often scanned for automated analysis, further processing, and archiving. Due to the nature of such documents, degraded document images are often hard to read, have low contrast, and are corrupted by various artifacts. We describe a novel approach for the enhancement of such documents based on probabilistic models which increases the contrast, and thus, readability of such documents under various degradations. The enhancement produced by the proposed approach can be viewed under different viewing conditions if desired. The proposed approach was evaluated qualitatively and compared to standard enhancement techniques on a subset of historical documents obtained from the Yad Vashem Holocaust museum. In addition, quantitative performance was evaluated based on synthetically generated data corrupted under various degradation models. Preliminary results demonstrate the effectiveness of the proposed approach.

Keywords: degraded document image enhancement, historical document image enhancement, document image analysis, document degradation models, image enhancement, image analysis

1. INTRODUCTION

Degraded documents are archived and preserved in large quantities worldwide. Electronic scanning is a common approach in handling such documents in a manner which facilitates public access to them. Such document images are often hard to read, have low contrast, and are corrupted by various artifacts. Thus, given an image of a faded, washed out, damaged, crumpled or otherwise difficult to read document, one with mixed handwriting, typed or printed material, with possible pictures, tables or diagrams, it is necessary to enhance its readability and comprehensibility. Documents might have multiple languages in a single page and contain both handwritten and machine printed text. Machine printed text might have been produced using various technologies with variable quality. The approach described herein is concerned with automatic enhancement of such documents and is based on several steps: the input image is segmented into foreground and background, the foreground image is enhanced, the original image is enhanced, and the two enhanced images are blended using a linear blending scheme. The use of the original image in addition to the foreground channel allows for foreground enhancement while preserving qualities of the original image. In addition, it allows for compensation for errors that might occur in the foreground separation.

The enhancement process we propose produces a document image that can be viewed in different ways using two interactive parameters with simple and intuitive interpretation. The first parameter controls the decision threshold used in the foreground segmentation whereas the second parameter controls the blending weight of the two channels. Using the decision threshold the user may increase or decrease the sensitivity of the foreground segmentation process. Using the blending factor the user can control the level of enhancement: on one end of the scale the original document image is presented without any enhancements, whereas on the other end, the enhanced foreground is displayed by itself. Note that the application of these two adjustable thresholds is immediate once the document image has been processed. The adjustment of the parameters is not necessary and is provided to enable different views of the document as deemed necessary by the user. The overall system architecture is depicted in Figure 1.

The remainder of the paper describes the proposed approach in greater detail. In Section 2, we describe related work and other approaches that have been employed to document image enhancement. In Section 3, we provide a detailed description of the proposed approach. Preliminary results of both qualitative and quantitative evaluation of the proposed approach obtained using actual and synthesized data are presented in Section 4. We concludes the paper in Section 5.
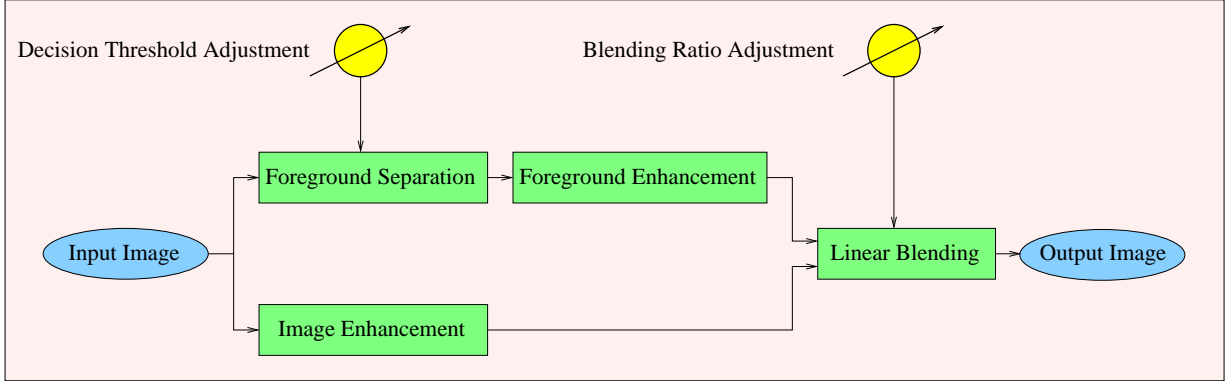
Figure 1. Description of the document enhancement system architecture (see text). After initial processing, the user can use two adjustable thresholds to control both the foreground separation decision threshold and the blending level.

2. RELATED WORK

The problem of linear image enhancement is well studied. When the degradation model is linear with additive Gaussian noise, optimal solution (in the sense of MSE) to the enhancement problem is available through the Weiner filter by estimating the optical transfer function of the degradation process and the noise characteristics. Common degradations in document images are in many cases non-linear, and so, require specialized treatment. A thorough discussion and evaluation of document degradation models is provided by Kanungo et al..[1]

A thorough review of segmentation techniques for grayscale document images is provided in.[2,3] The comparison is based on OCR accuracy rate after segmentation. Using this metric, it is concluded that local segmentation techniques have higher performance. In particular, a method which is based on a local minimum and maximum values in each sub-window[4] is shown to be one of the most efficient and effective techniques. Approaches for text segmentation in color images are described in by Perroud et al.[5] and Loo and Tan.[6] A component based approach for foreground-background separation in low-quality color document images is described by Garain et al..[7,8] In this approach, connected components are labeled and organized in tree structures. Nodes in the tree are then segmented using K-means. The performance of this approach is evaluated by measuring the improvement in word and line segmentation algorithms before and after enhancement.

A commercial-grade algorithm for foreground-background separation is part of the DjVu system.[9,10] The DjVu system is interested primarily in document image compression and uses foreground-background separation to store foreground information with higher detail. The DjVu algorithm uses the K-means algorithm to perform clustering of foreground and background pixels in each sub-window. To assure consistency between neighboring sub-windows, and avoid problems associated with selecting a specific window size, the clustering is performed at multiple scales of the image. The cluster centers at a given scale are initialized based on the results of the previous (coarser) scale in which a larger window is used. The cluster centers at each iteration of the K-means algorithm are computed based on a weighted sum of the cluster centers found in the previous (coarser) scale and the average of the cluster members in the current scale. Finally, various heuristics are used to correct the foreground segmentation results. The DjVu foreground-background separation algorithm is available through a web interface (http://any2djvu.djvuzone.org/). A simpler and more efficient approach for document image compression based on background-foreground separation is described by Simard et al..[11]

An approach for the enhancement of faxed documents that is based on assumptions of character regularity is described by Hobby and Ho.[12] In this approach, bitmaps of identical symbols present on the same page are clustered and averaged. The averaged symbols are then used to enhance the image. This is a particularly effective technique for enhancing low resolution scans by combining multiple low-resolution instances of the same symbol to obtain a higher resolution of it. Binarization of historical documents based on adaptive threshold segmentation and various pre- and post-processing steps is described by Gatos et al..[13] In this approach, a background surface is estimated and used to segment the image. An iterative approach for segmenting degraded

document images is described by Kavallieratou et al..[14] There, a global thresholding technique is used to obtain an initial segmentation. Areas with likely incorrect segmentation are detected, and a local thresholding is applied in them. This approach is efficient in that local thresholds are computed only at selected locations. It is also noted that general-purpose segmentation techniques provided better performance on historical documents as compared with document-specific segmentation techniques.

3. THE PROPOSED APPROACH

The proposed approach for document image enhancement is composed of several steps including foreground segmentation, foreground enhancement, image enhancement, and linear blending (see Figure 1). The foreground segmentation step is clearly the most difficult and is addressed in detail later in this section. Once the foreground has been separated, its enhancement is performed by creating a binary mask and reducing the intensity of pixels that are not masked out. When the reduction is severe (e.g., to 0), the resulting foreground channel I_f can then be convolved with a Gaussian to produce a slightly smoother result. The image enhancement step involves simple filtering operations (such as the median filter) intended to improve the overall quality of the image, and is applied to the complete image channel I_i. The linear blending step then sets the output image to $I_o = (1 - \lambda)I_i + \lambda I_f$ thus allowing for a smooth transition between the original image ($\lambda = 0$) and the enhanced foreground channel ($\lambda = 1$). Without user interaction the blending coefficient λ is set to 0.5.

Perhaps the simplest approach for two-class foreground segmentation is the weighted-mean approach. In this approach, the image is traversed by a sliding window. A local threshold is determined for each sub-window and applied to perform the segmentation of the pixels contained in it. The threshold $\tau_{i,j}$ in the (i, j) sub-window is computed based on a weighted sum of the intensity values which is then decreased by a constant value d provided as a parameter. The weights in the weighted sum are given by a Gaussian function $G_\sigma(i, j)$ with a mean vector of (i, j) and a covariance matrix of $\text{diag}(\sigma, \sigma)$. That is, $\tau_{i,j} = \sum_k \sum_l I(i + k, j + l) \cdot G_\sigma(i + k, j + l) - d$. Given the properties of Gaussian functions, the variance value σ is normally taken to be one third of the window size.

The weighted-mean local segmentation as described above assumes a roughly equal number of foreground and background pixels in each sub-window, and so, will fail in nearly uniform regions in which only foreground pixels are present by producing a threshold which is too high. A possible solution to this problem is to examine the distribution of the obtained threshold values, identify outliers, and replace outlier threshold values with an interpolated value obtained from neighboring threshold values using bilinear interpolation. A different solution to this problem, termed the min-max approach herein, may be obtained by computing the minimum $I_{\min}(i, j)$ and maximum $I_{\max}(i, j)$ intensity values in the (i, j) sub-window and computing the local threshold value as a value between them. That is, $\tau_{i,j} = I_{\min}(i, j) + \rho(I_{\max}(i, j) - I_{\min}(i, j))$, where ρ is a decision parameter provided by the user. The default value of ρ is taken as 0.5. The value of $\tau_{i,j}$ is computed as above only if $(I_{\max}(i, j) - I_{\min}(i, j)) > \alpha$ where α is a preset threshold. If this condition is not satisfied, the sub-window does not contain both foreground and background pixels, and so, in such a case, the threshold $\tau_{i,j}$ is set to 256. As the minimum and maximum values are sensitive to outliers, it is possible to either smooth the image using a 3×3 median filter prior to computing the minimum and maximum values, or use the 10th and 90th percentile pixels instead on the minimum and maximum values, respectively. To remove small noise artifacts from the segmentation results, a filter that removes small connected components is applied to it. This is handled with care, as many languages and scripts contain diacritics whose size, in pixel count, is very small. The min-max algorithm as described here is an extension of an existing segmentation technique[4] which was found to be very efficient and effective in a comparative evaluation of multiple segmentation techniques.[3]

The min-max algorithm as described above works well to some extent. However, it involves several parameters which are set manually. Manual setting of parameters is not appropriate for large and diverse collections in which the parameters need to be readjusted. Estimating the parameters automatically is a difficult problem that needs to be addressed. Moreover, the min-max algorithm is based on intensity features and does not extend directly to include additional features such as edge features. To overcome these problems, we employed the expectation maximization algorithm (EM) which is a general algorithm that can handle in a generic way multiple features simultaneously.

In the proposed approach, mixture models[15] are used to model the joint distribution of features. A mixture model with M components $p_i(x|\theta_i)$ is given by: $p(x|\Theta) = \sum_{i=1}^{M} \alpha_i p_i(x|\theta_i)$, where the mixing coefficients α_i

satisfy $\sum_{i=1}^{M} \alpha_i = 1$ and $\Theta = (\alpha_1, \ldots, \alpha_M, \theta_1, \ldots, \theta_M)$. In the absence of any specific information regarding the parametric form of the mixture components, a common choice is to model them through multivariate Gaussians. Alternatively, a relatively more efficient form computationally which does not include exponents, are clamped multivariate Cauchy distributions given by:

$$p_i(x|\theta_i) = \frac{\pi^{-d}|\Sigma_i|^{-1/2}}{(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i) + 1} \tag{1}$$

where μ_i and Σ_i are the mean vector and covariance matrix, respectively, d is the dimensionality of the data, and the parameter vector θ_i is composed of the elements of μ_i and Σ_i. Since the variance of the Cauchy distribution is infinite, the elements of Σ_i should be interpreted in the sense of full width at half maximum.

Let $\mathcal{X} = \{x_i\}_{i=1}^{N}$ be a set of identically and independently distributed observations distributed according to the mixture density function $p(x|\Theta)$. Based on the independence assumption, the incomplete-data log-likelihood is given by:

$$l(\Theta|\mathcal{X}) = \log \prod_{i=1}^{N} p(x_i|\Theta) = \sum_{i=1}^{N} \log \left(\sum_{j=1}^{M} \alpha_j p_j(x_i|\theta_j) \right) \tag{2}$$

The unknown parameter vector Θ can be obtained by maximizing $l(\Theta|\mathcal{X})$.

Following the well known expectation maximization (EM) algorithm,[16] it is possible to simplify the maximization of the incomplete-data log-likelihood by assuming a hidden feature describing the unknown component in the mixture from which each observation was drawn. Let y_i be the hidden feature corresponding to the observation x_i. Taking the expectation over the hidden features, the expected complete-data log-likelihood can be estimated iteratively by:

$$Q(\Theta, \Theta^{(s)}) = \sum_{l=1}^{M} \sum_{i=1}^{N} \log(\alpha_l) p(l|x_i, \Theta^{(s)}) + \sum_{l=1}^{M} \sum_{i=1}^{N} \log(p_l(x_i|\theta_l)) p(l|x_i, \Theta^{(s)}) \tag{3}$$

where $\Theta^{(s)}$ is the estimate of Θ at iteration s and the posterior component probability $p(l|x_i, \Theta^{(s)})$ is given by:

$$p(l|x_i, \Theta^{(s)}) = \frac{\alpha_l^{(s)} p_l(x_i|\theta_l^{(s)})}{\sum_{k=1}^{M} \alpha_k^{(s)} p_k(x_i|\theta_k^{(s)})} \tag{4}$$

Explicit equations for the mixing coefficients and the component distribution parameters can then be derived. Features tested in our experimental evaluation of foreground segmentation include both color and edge features for both two-class and three-class segmentation.

The EM algorithm is optimal in the sense of maximizing the incomplete-data log-likelihood. Yet the performance of this algorithm can be improved by considering local neighborhoods separately. This is due to the fact that degradations can vary across different areas of a document. The selection of a suitable window size can affect the segmentation results. If the window size is too small it might not have a sufficient number of pixels belonging to both the foreground and background classes. Conversely, if the window size is too large it might contain several different degradations and not perform optimally. To select the window size adaptively, the proposed approach begins with an initial window size that is narrowed down iteratively as long as the distance between the estimated means is sufficiently large and as long as the change in the means is sufficiently small with respect to their previous value. The initial window size is estimated automatically based on the estimated average distance between text lines. An example of adaptive window determination is presented in Figure 2. Figure 2-(a) shows the original image whereas Figure 2-(b) shows the resulting adaptive window size estimation.

4. RESULTS AND DISCUSSION

The performance of the different approaches was evaluated qualitatively using a subset of historical documents obtained from the Yad Vashem Holocaust museum.[17] The test collection contains 867 pages written in several different languages which were produced mainly by a typewriter. The documents contain numerous handwritten
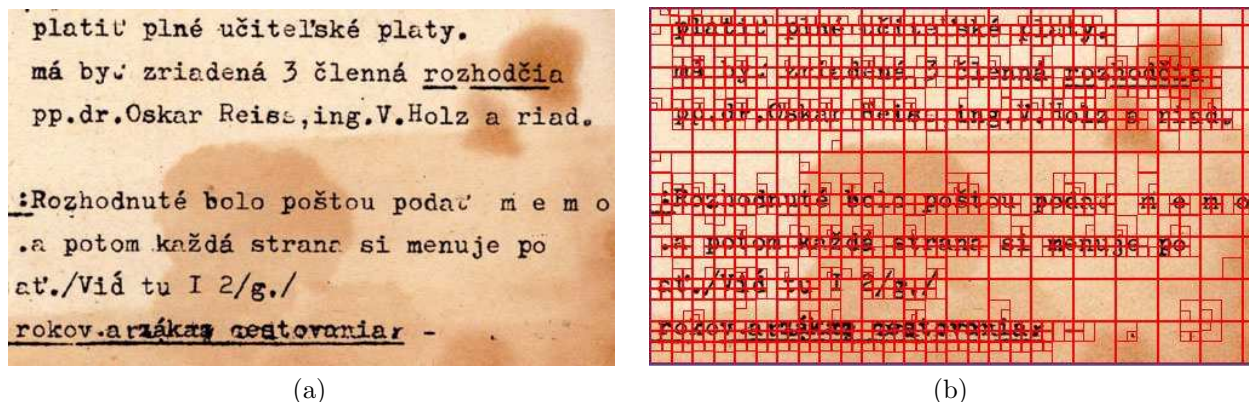
Figure 2. Adaptive window size determination. **(a)** The original image. **(b)** The resulting adaptive window size estimation.

comments as well as logos and signatures. They are mainly 60-70 years old and exhibit different levels of degradation due to various sources. Sample results of this evaluation are presented in Figure 3. Figure 3-(a) shows the original image, Figures 3-(b) – 3-(f) show the enhancement results obtained using the commercial grade DjVu algorithm, the min-max algorithm, the two-class color EM, the three-class color EM, and the three-class color and edge EM, respectively. As can be observed, the two-class color EM and the three-class color and edge EM provide the best visual enhancement in the case of faded characters and non-uniform background. Note that different algorithms perform differently on different documents, and so, if the perceptual quality can be assessed automatically, the best result can be selected.
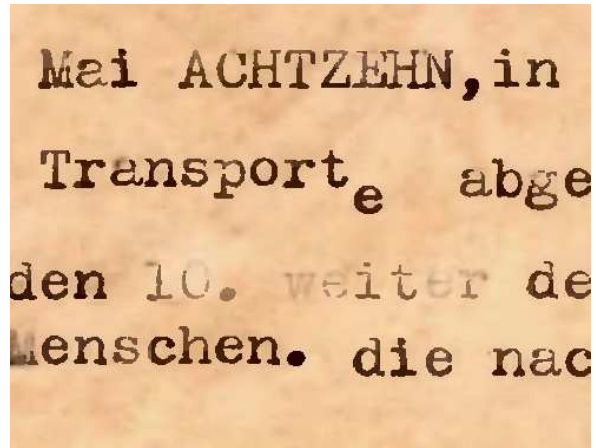
The different enhancement methods were evaluated on pure degradations including local-brightness degradation, blurring degradation, noise degradation, and texture-blending degradation. The local-brightness degradation simulates effects such as uneven key pressure in typewriter produced documents, or faded ink in handwritten documents. This degradation was produced by randomly selecting rectangular windows in the image and increasing their brightness by adding a constant brightness and clamping the obtained intensity. The blurring degradation simulates effects such as fading or writing with imprecise writing instruments, and was produced by convolving the image with a Gaussian. The noise degradation simulates effects such as imperfect typing and dirt, and was produced by randomly flipping the values of pixels in the image. Finally, the texture-blending degradation simulates effects such as textured paper or stained paper, and was produced by linearly blending the document with a texture image. Illustration of these degradations are provided in Figure 4.

The evaluation includes comparison of the commercial grade DjVu foreground segmentation algorithm, the min-max algorithm as described before, a two-class EM algorithm based on texture features, and a two-class EM algorithm based on texture and edge features. The evaluation scheme assumes the knowledge of the original document without any degradations in which two classes (foreground and background) are available. Consequently, the three-class segmentation techniques were excluded from this evaluation. Examples of the results obtained by the different approaches in enhancing local brightness, and blurring, is provided in Figures 5 and 6, respectively. As can be observed, different algorithms perform better on different kinds of degradations. Specifically, the EM based algorithms perform the best on local brightness and blurring degradations, the min-max algorithm performs better on noise degradations, and the DjVu and the color-based EM algorithms perform better on texture blending degradations. Note that the EM algorithm does not perform well on the salt-and-pepper noise degradation due to the fact that it does not consider spatial smoothing constraints.

To quantify the performance of each of the compared techniques, the degraded images were segmented using the different approaches, and the segmented results were compared to the known ground truth image. The true positive (TP), false negative (FN), and false positive (FP) rates, were measured and then converted to precision (TP/(TP+FP)) and recall (TP/(TP+FN)) rates. Precision-recall graphs were then generated for each method by varying the decision threshold of each method. The DjVu algorithm does not provide access to a decision
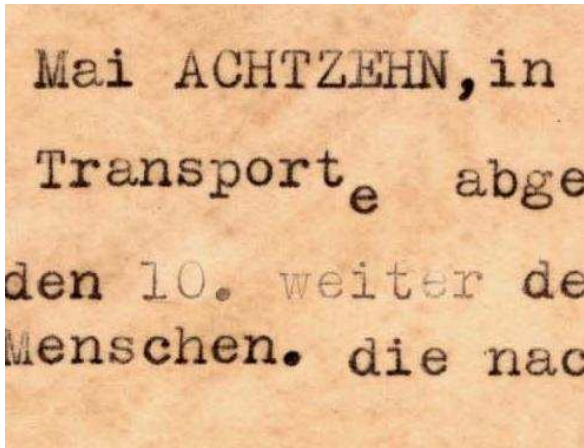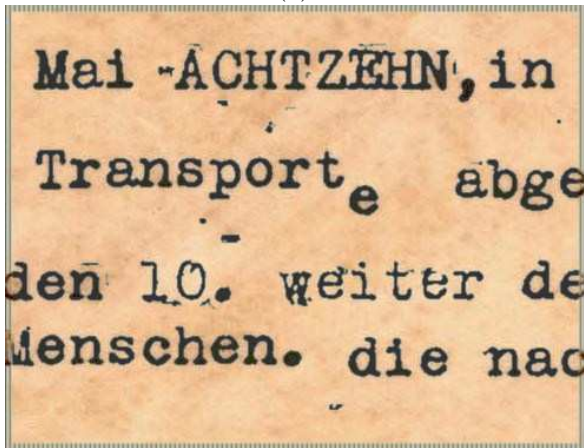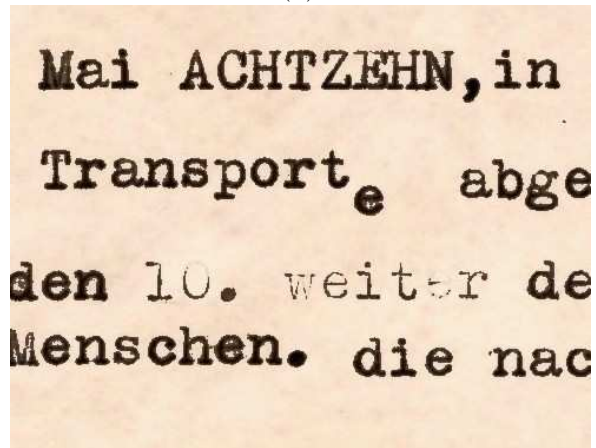
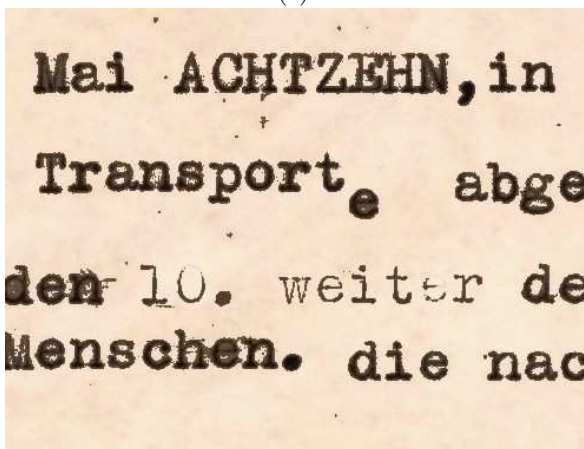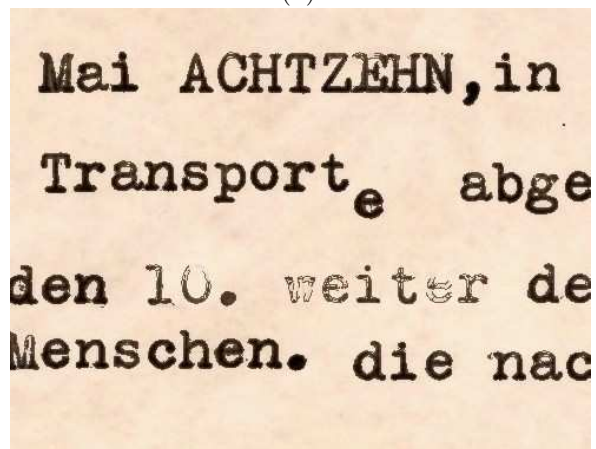**Figure 3.** Qualitative comparison of different enhancement results. **(a)** The original image. **(b)** DjVu enhancement. **(c)** Min-max enhancement. **(d)** Two-class color EM enhancement. **(e)** Three class color EM enhancement. **(f)** Three class color and edge EM enhancement.
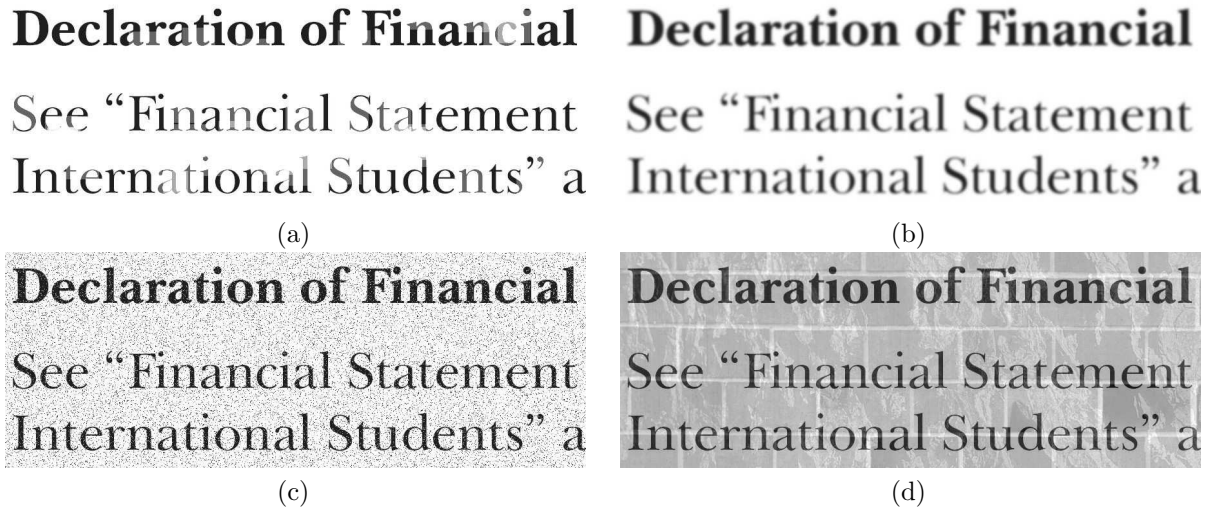
Figure 4. Illustration of pure synthetic degradations. **(a)** Local brightness degradation. **(b)** Blurring degradation. **(c)** Noise degradation. **(d)** Texture blending degradation.
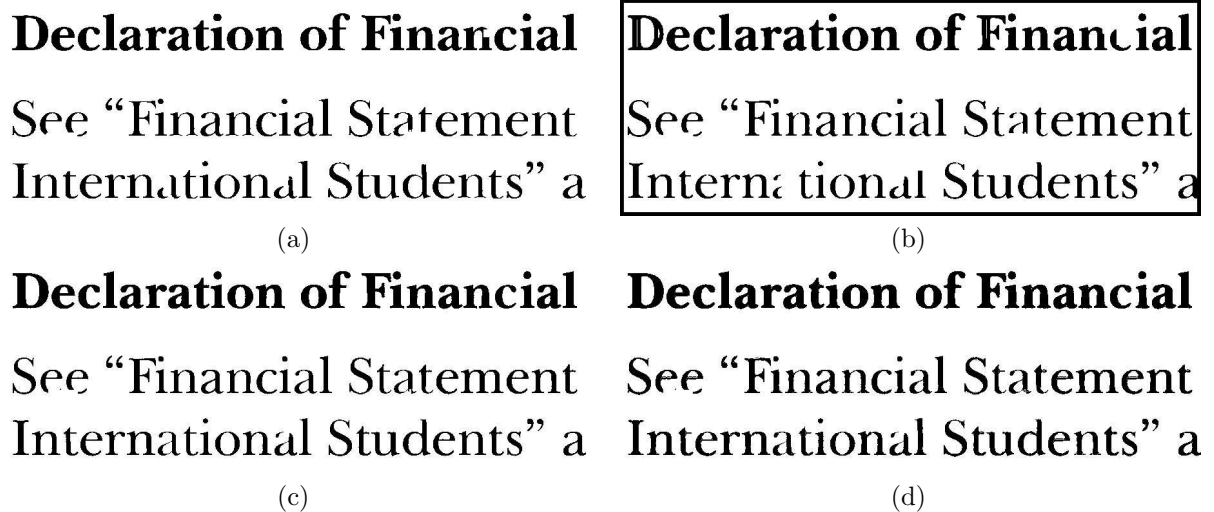

Figure 5. Example of enhancement of a local-brightness degradation using the different approaches. **(a)** DjVu foreground segmentation. **(b)** Min-max segmentation. **(c)** Two class color-based EM. **(d)** Two class color- and edge-based EM.

threshold and was not included in this evaluation. The results of this evaluation for different degradation models are presented in Figures 7 and 8. Figure 7 shows precision and recall curves for foreground segmentation results using the min-max algorithm (MM), the two-class color-based EM algorithm (EM2C), and the two-class color- and edge-based EM algorithm (EM2CE). The evaluation is performed on local-brightness (top) and blurring (bottom) degradations. Figure 8 displays a similar evaluation which is performed on noise and texture blending degradations. As can be observed, the EM algorithm attains a higher precision for a given recall rate, and so, performs better in most cases. The improved performance of EM is also evident when considering the area under the precision-recall curve which is larger for EM. Note that for the noise degradation, the min-max algorithm produces better results in some cases. This is due to the fact that our current implementation of the EM algorithm does not consider spatial smoothing constraints. Concerning the behavior of precision and recall as
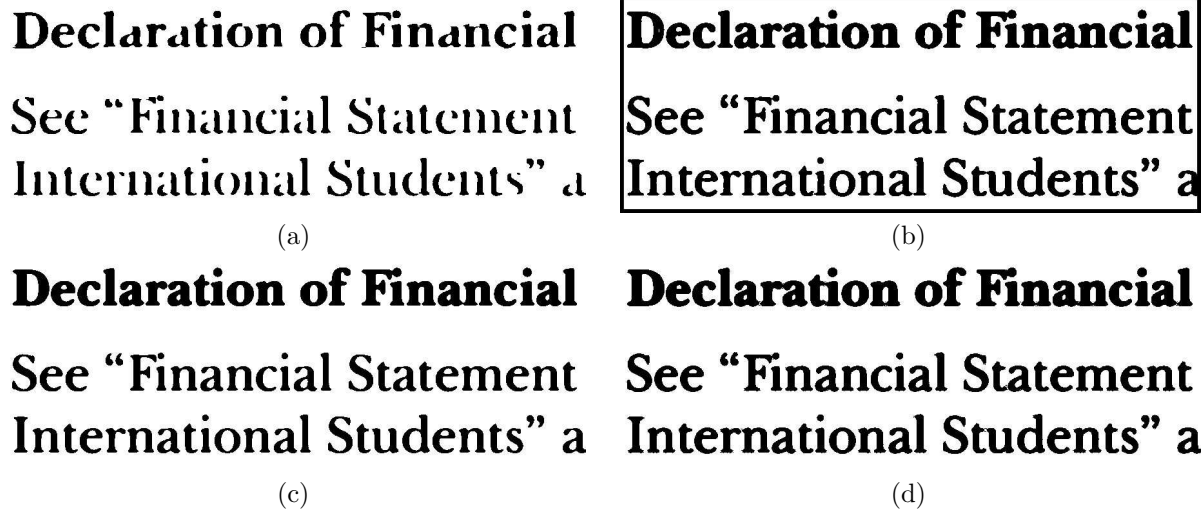
Declaration of Financial

See "Financial Statement

International Students" a

(a)

Declaration of Financial

See "Financial Statement

International Students" a

(b)

Declaration of Financial

See "Financial Statement

International Students" a

(c)

Declaration of Financial

See "Financial Statement

International Students" a

(d)

Figure 6. Example of enhancement of a blurring degradation using the different approaches. **(a)** DjVu foreground segmentation. **(b)** Min-max segmentation. **(c)** Two class color-based EM. **(d)** Two class color- and edge-based EM.

a function of the decision threshold value, it is possible to observe that increments to the decision threshold always result in increased recall rates. In the case of the min-max algorithm, these increased recall rates are associated with increased precision rates, whereas in the case of the EM algorithms these increased recall rates are associated with either increased or decreased precision rates. This is due to the more complex nature of the decision surface that is produced by the EM algorithms. Note that in contrast to specificity measurements in ROC curves, precision measurements are not biased by the presence of large amounts of negatives (background pixels), and so, are more accurate in evaluating performance of degraded image documents.

5. SUMMARY

We described a novel approach for the enhancement of degraded document images with multiple possible degradations. The degradations are handled in a generic way through either a min-max or probabilistic model (EM). The enhancement results produced allow for a continuous set of views which are controlled by two interactive parameters. One parameter controls the decision threshold of the foreground separation process, and the second controls the blending factor of the display. While it is not necessary to change the viewing parameters, the incorporation of such viewing parameters allow for improved visualization. The proposed approach was evaluated both qualitatively and quantitatively and compared to the commercial grade DjVu algorithm using various degradation models. Future work will address the incorporation of additional features, development of quantitative perceptual quality measures, and development of methods for selecting particular enhancement models for individual cases.

Acknowledgments

REFERENCES

1. T. Kanungo, R. Haralick, H. Baird, W. Stuezle, and D. Madigan, "Statistical, nonparametric methodology for document degradation model validation," *IEEE Trans. Pattern Analysis and Machine Intelligence* **22**(11), pp. 1209–1223, 2000.
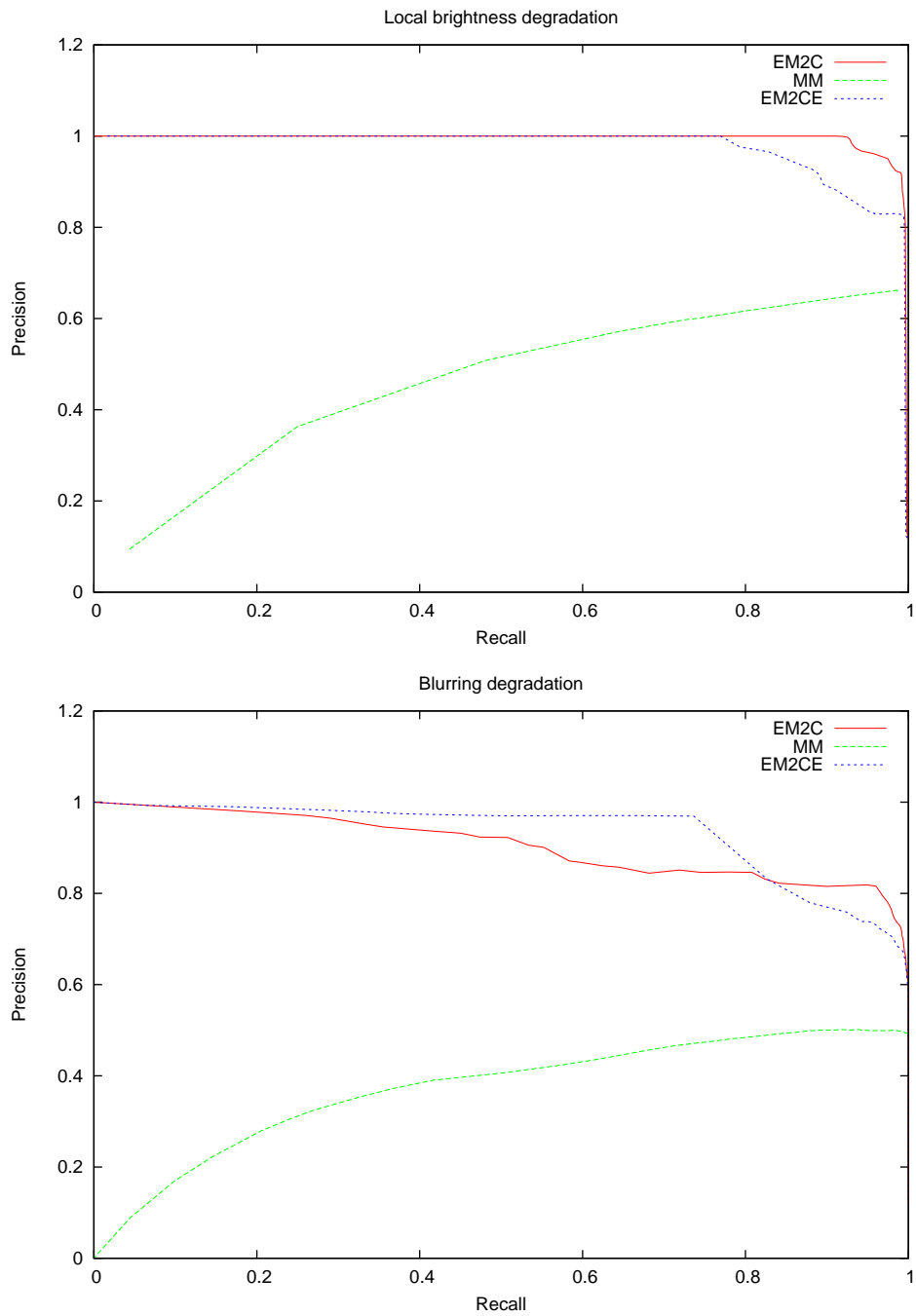
Figure 7. Precision and recall graphs for foreground segmentation results using the min-max algorithm (MM), the two-class color-based EM algorithm (EM2C), and the two-class color- and edge-based EM algorithm (EM2CE). Evaluation on local-brightness (top) and blurring (bottom) degradations.
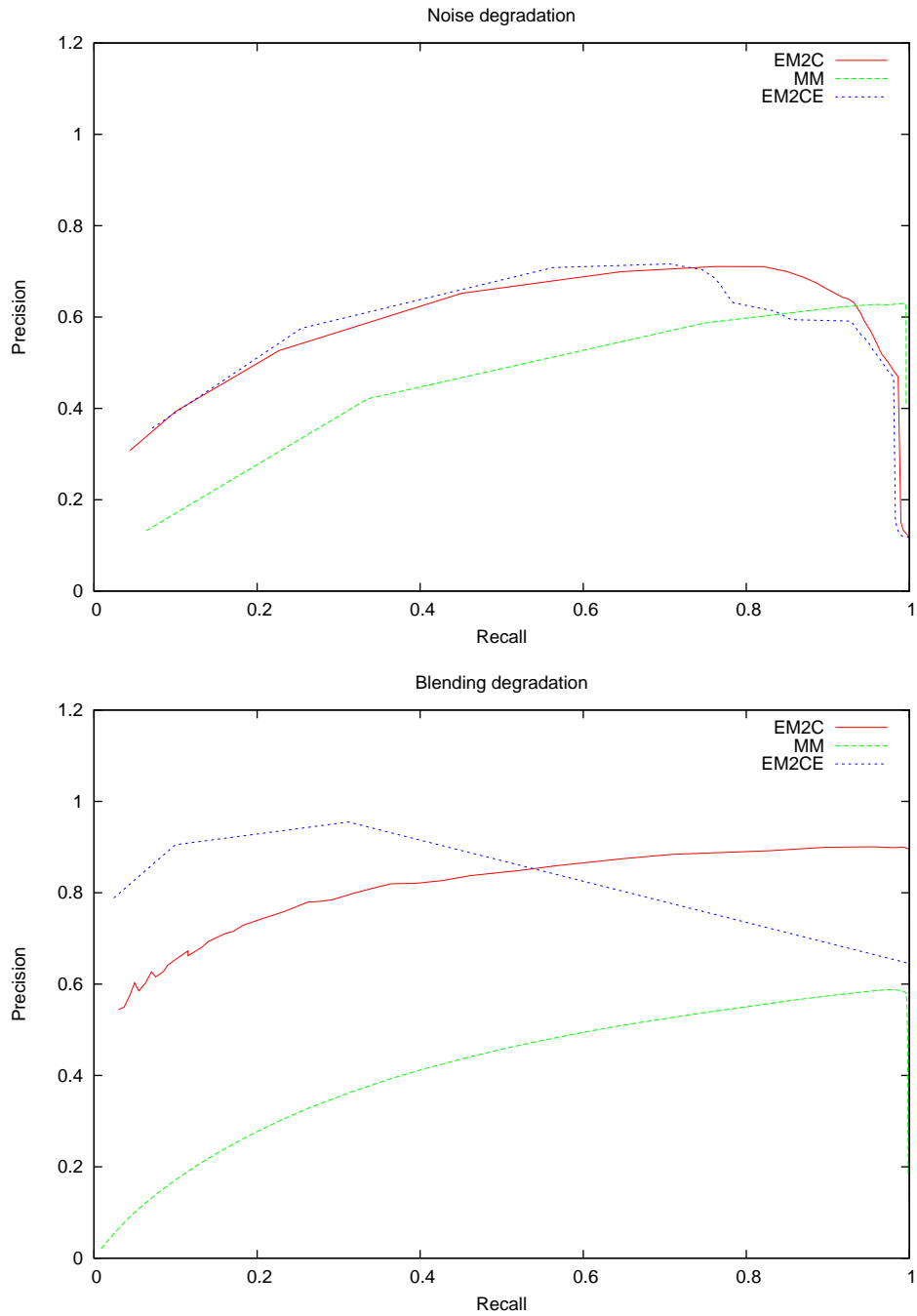
Figure 8. Precision and recall graphs for foreground segmentation results using the min-max algorithm (MM), the two-class color-based EM algorithm (EM2C), and the two-class color- and edge-based EM algorithm (EM2CE). Evaluation on noise (top) and texture blending (bottom) degradations.

2. O. Trier and T. Taxt, "Evaluation of binarization methods for document images," *IEEE Trans. Pattern Analysis and Machine Intelligence* **17**(3), pp. 312–315, 1995.

3. O. Trier and A. Jain, "Goal-directed evaluation of binarization methods," *IEEE Trans. Pattern Analysis and Machine Intelligence* **17**(12), pp. 1191–1201, 1995.

4. J. Bernsen, "Dynamic thresholding of gray-level images," in *Proc. Int'l Conf. Pattern Recognition (ICPR)*, **2**, pp. 1251–1255, 1986.

5. T. Perroud, K. Sobottka, and H. Bunke, "Text extraction from color documents-clustering approaches in three and four dimensions," in *Proc. Int'l Conf. Document Analysis and Recognition (ICDAR)*, pp. 937–941, 2001.

6. P. K. Loo and C. L. Tan, "Adaptive region growing color segmentation for text using irregular pyramid," in *Proc. Int'l Workshop Document Analysis Systems (DAS)*, pp. 264–275, (Florence, Italy), 2004.

7. U. Garain, T. Paquet, and L. Heutte, "On foreground-background separation in low quality color document images," in *Proc. Int'l Conf. Document Analysis and Recognition (ICDAR)*, **2**, pp. 585–589, (Seoul, Korea), 2005.

8. U. Garain, T. Paquet, and L. Heutte, "On foreground - background separation in low quality document images," *Int'l J. Document Analysis and Recognition* **8**(1), pp. 47–63, 2006.

9. L. Bottou, P. Haffner, P. G. Howard, P. Simard, Y. Bengio, and Y. LeCun, "High quality document image compression with DjVu," *J. Electronic Imaging* **7**(3), pp. 410–425, 1998.

10. P. Haffner, L. Bottou, P. Howard, and Y. LeCun, "DjVu: Analyzing and compressing scanned documents for internet distribution," in *Proc. Int'l Conf. Document Analysis and Recognition (ICDAR)*, pp. 625–628, 1999.

11. P. Simard, H. Malvar, J. Rinker, and E. Renshaw, "A foreground-background separation algorithm for image compression," in *Proc. Data Compression Conference (DCC)*, pp. 498–507, 2004.

12. J. Hobby and T. K. Ho, "Enhancing degraded document images via bitmap clustering and averaging," in *Proc. Int'l Conf. Document Analysis and Recognition (ICDAR)*, **1**, pp. 394–400, (Ulm, Germany), 1997.

13. B. Gatos, I. Pratikakis, and S. J. Perantonis, "An adaptive binarization technique for low quality historical documents," in *Proc. Int'l Workshop Document Analysis Systems (DAS)*, pp. 102–113, (Florence, Italy), 2004.

14. E. Kavallieratou and E. Stamatatos, "Improving the quality of degraded document images," in *Proc. Int'l Conf. Document Image Analysis for Libraries (DIAL)*, (Lyon, France), 2006.

15. G. Agam and C. Wu, "Probabilistic modeling based vessel enhancement in thoracic CT scans," in *IEEE Intl. Conf. on Computer Vision and Pattern Recognition (CVPR)*, **2**, pp. 684–689, (San Diego, CA), 2005.

16. C. Fraley and A. E. Raftery, "Model-based clustering, discriminant analysis, and density estimation," *Journal of the American Statistical Association* **97**, pp. 611–631, 2002.

17. "The diaries of Rabbi Dr. Avraham Abba Frieder." http://ir.iit.edu/collections/.