01 Jan 2008

Efficient MRF approach to document image enhancement

Tayo Obafemi-Ajayi
Missouri University of Science and Technology, towd2@mst.edu

Gady Agam

Ophir Frieder

Recommended Citation

T. Obafemi-Ajayi et al., "Efficient MRF approach to document image enhancement," *Proceedings - International Conference on Pattern Recognition*, article no. 4761557, Institute of Electrical and Electronics Engineers, Jan 2008.
The definitive version is available at https://doi.org/10.1109/icpr.2008.4761557

Efficient MRF Approach to Document Image Enhancement

Tayo Obafemi-Ajayi, Gady Agam, Ophir Frieder
Illinois Institute of Technology
ihimtay@iit.edu, agam@iit.edu, ophir@ir.iit.edu

Abstract

Markov random field (MRF) based approaches have been shown to perform well in a wide range of applications. Due to the iterative nature of the algorithm, the computational cost of such applications is normally high. In the context of document image analysis, where numerous documents have to be processed, this computational cost may become prohibitive. We describe a novel approach to document image enhancement using MRF. We show that by using domain specific knowledge, we are able to substantially improve computational performance by an order of magnitude. Moreover, in contrast to known techniques where patch initialization is arbitrary, in the proposed approach patch initialization is data consistent and so results in improved effectiveness. Experimental results comparing the proposed approach to known techniques using historical documents from the Frieder Collection are provided.

1. Introduction

Markov Random Fields (MRF) based approaches have attained significant success in the field of image restoration [6]. Thus, there is a growing interest in applying MRF techniques to document image restoration. Spatial methods for text and handwritten images utilizing the MRF probabilistic technique (such as [9], [7], [10], [5]) have a key advantage over heuristic algorithms due to their ability to learn the probabilistic dependency of neighboring pixels or image patches (the prior probability) from training data by modeling their relationship using Markov networks. However, given the iterative nature of the algorithm, the computational cost of such methods is normally high. We describe a novel MRF approach to document image enhancement that utilizes domain specific knowledge to substantially improve computational performance by an order of magnitude, resulting in a more efficient algorithm.

Recent work done on text and handwritten images based on MRF techniques include the following. Wolf et al. in [9] applied the MRF concept to the binarization of low quality text from multimedia documents. From their results, their method does not improve existing techniques significantly, partly due to the sensitivity of their method to model parameters (the noise variance) and to the energy function. Obtaining a good observation model is essential to producing good results when using an MRF based approach. Gupta et al. [7] employ MRF, based on partial message propagation, to the recognition and restoration of blurred images of license plate digits. To segment strokes of Chinese characters, Zeng et al. [10] formulate the stroke segmentation as an optimal labeling problem using a *maximum a posteriori* (MAP) MRF model. This MAP-MRF framework improves the accuracy of stroke segmentation by modeling the relationships among neighboring sites statistically.

Cao et al. [5] apply MRF model as a binarization technique in the context of preprocessing handwritten carbon form document images. Their algorithm uses a collection of standard patches (obtained by clustering all patches of binarized images in the training set) to represent each patch of the binarized image from the test set. This reduces the domain of the prior model to a very limited size. Rather than using an image/scene pair for learning the observation model, it is learned on-the-fly from the local histogram of the test image. The performance of the Cao et al. algorithm is biased by the quality of the initial segmentation algorithm. When applied on degraded document images with very poor initial segmentation, it performs poorly since the density estimates are much less accurate.

In this paper, we present an efficient and improved MRF algorithm in the context of enhancement of degraded historical typewritten document images. Historical documents generally suffer large degradations which hinder their readability. Antonacopoulos et al. expound on the unique challenges facing enhancement of typewritten documents in [2] such as very poor contrast due to the age of the paper and ink used, blurred or faint text due to uneven typewriter key pressure or faded ink. In contrast to known techniques where the probabilistic density of the observation model is learned-on-the-fly, the proposed MRF approach employs an offline estimation process. Moreover, differing from the current state of the art MRF techniques where patch initialization is arbitrary, our patch initialization is data consistent and so results in improved effectiveness. Experimental results comparing the proposed approach to known techniques using historical documents from the

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

Frieder Collection [1] are provided.

2. Proposed MRF-based Enhancement

In contrast to natural scene images, document images have inherent structure that can be exploited for improved performance. By deducing the underlying structure it is possible to decouple the processing of different image parts thus resulting in multiple simplified and disjoint models. This can then be translated to improved computational performance as well as improved effectiveness. The underlying structure that has to be inferred is application dependent. In the context of document image enhancement, where the focus is the enhancement of characters, character segmentation can be used to decouple the models. Character segmentation in machine printed text is relatively simple when assuming regularity in the document and is routinely performed in production quality document imaging systems. As shown later, using decoupling based on character segmentation we are able to improve computational performance by an order of magnitude.

Generated MRF models are often based on image patches which are generated in a regular fashion and so have random alignment with image contents. This produces an unnecessary increase in the number of different patches which in turn increases the complexity of the problem. The consequence of this unnecessary increase is reduced and inconsistent performance which depends on random alignment. By using character segmentation we are able to generate a more consistent patch initialization which is data dependent and so improve effectiveness.

2.1 The MRF model

The goal is to infer an enhanced binary representation x (a clean image) of a degraded document image y. We partition x and y into non-overlapping square (of size 5×5) patches $x_1, x_2, ..., x_N$ and $y_1, y_2, ..., y_N$ such that y_i is the observation of x_i for $1 \leq i \leq N$ [5], [6]. The Markov assumption [6] is that each scene (in the binarized image) patch is statistically dependent on the corresponding image (in the degraded image) patch and to its spatial neighbors. The spatial neighbors of x_i are its four neighboring patches in both horizontal and vertical directions. Using the MRF model, we can find the best estimate for x as the maximum a posteriori (MAP) of the posterior probability given as $P(x|y) = cP(x, y)$ (the normalization constant $c = \frac{1}{P(y)}$ is constant over x).

Solving an MRF involves both learning and inference phases. During the inference phase, the MRF model uses statistical properties gathered from the labeled training data (during the learning phase) to form "best guess" estimates of x using equation 1. The local evidence is propagated via belief propagation (BP) to the neighboring nodes to determine the maximum a posteriori probability (MAP) of the binarized image estimate. This estimate is given by:

$$\hat{x}_{j\ MAP} = \underset{x_j}{argmax}\ \mathrm{P}(x_j)P(y_j|x_j)\prod_k M_j^k \qquad (1)$$

where k runs over all the four neighboring nodes of node j in the binarized image x and M_j^k, the message from node k to node j. The message M_j^k is computed by:

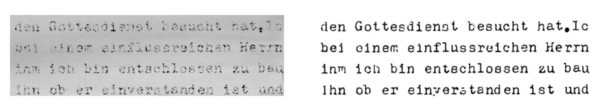
$$M_j^k = \underset{x_k}{max}\ \mathrm{P}(x_k|x_j)P(y_k|x_k)\prod_{l \neq j} \tilde{M}_k^l \qquad (2)$$

where $\tilde{M}_k^l$ is M_j^k from the previous iteration, initialized as column vectors of 1's.

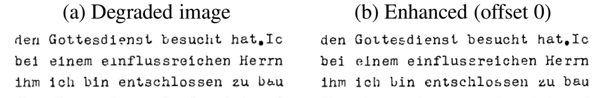
Equation (1) is the factorized form of the product of compatibility functions of the neighboring nodes using MAP estimation [6]. Belief propagation operates by passing messages between nodes using the message update equation (2) until convergence.

2.2 Patch alignment

For any document image, depending on how we align the starting point of the patches, we can obtain different set of patches from the same image. Since the MRF algorithm divides the image into a set of non-overlapping patches, a question that arises is how do we align the starting point of the patches? If the algorithm is not significantly sensitivity to the alignment, then the results obtained should not differ if we vary the starting point. We performed preliminary experiments to verify this by simply varying the offset of the initialization point of the patches from 0 to 4. Both the qualitative and quantitative results obtained, as shown in Fig. 1, demonstrate that how we align our patches does affect the performance of the algorithm.



(a) Degraded image (b) Enhanced (offset 0)



(c) Enhanced (offset 2) (d) Enhanced (offset 4)

Figure 1: **Sensitivity of MRF to patch alignment.**

A degraded typewritten document image can be viewed as collection of character images. Our goal of enhancing the entire image is combination of the enhancement of all the characters in the image. We propose to break up the image intelligently into a set of sub-images, apply the MRF algorithm to each small image, and then combine the individual results for the final solution. The algorithm will converge much faster since the set of nodes for the message passing equation (2) is much less. If each small image potentially contains a single character, then the patches will be locally aligned on the characters rather than an arbitrary alignment process. Another advantage of apply character segmentation is that we can ignore areas in the document image that that contain no foreground information. To attain this, we preprocess the document image by applying the character segmentation algorithm employed in [3], to partition the image into a set of character bounding boxes, and enhance each individual box by applying the MRF algorithm.

2.3 Model learning

To use Equations (1) and (2) during the MRF inference phase, we need to model three probability densities [6]: the prior probability of each binarized image patch x_j i.e. $P(x_j)$, the conditional probability of a binarized image patch x_k and a neighboring patch x_j i.e. $P(x_k|x_j)$, and the conditional probability of an observation y_j given its estimate x_j i.e $P(y_j|x_j)$.

We learn the probability densities $P(x_j)$ and $P(x_k|x_j)$ from our training data set. The data set is a labeled portion of the subset of historical documents drawn from the Frieder collection [1]. The labeling was done by a human expert using an interactive document enhancement system [3]. Similar to handwriting images [5], machine printed (including typewritten) images under same resolution and similar font types can be decomposed into patches that appear frequently. Hence, using binary image patches of size 5×5, we cluster approximately 9 million patch images to generate a set of 150 standard patches $\{u_k\}$, as shown in Fig. 2.

To obtain the patch images for the clustering, we also preprocess the ground truth images using character segmentation to obtain a set of character bounding boxes. This ensures correlation of neighboring patches withing the bounding box. The prior probability $P(x_j)$ is the probability of each standard patch i.e. $x_j \in \mu = \{\mu_k\}$ estimated as relative occurrences [5]. $P(x_j, x_k)$ are also estimated in both horizontal and vertical directions as $P(x_j, x_k)(x_j, x_k \in \mu = \{\mu_k\})$ in a similar way.

By quantifying each patch as a vector, the conditional probability of the observation given a binary estimate is computed as [5]:

$$P(y_j|x_j) = \prod_{x_{j,j'}=0} p_b(y_{j,j'}) \prod_{x_{j,j'}=1} p_f(y_{j,j'}) \qquad (3)$$

where $x_{j,j'}$ runs over all elements in patch vector x_j and $y_{j,j'}$ is the observation of $x_{j,j'}$, $p_f(y)$ and $p_b(y)$ are

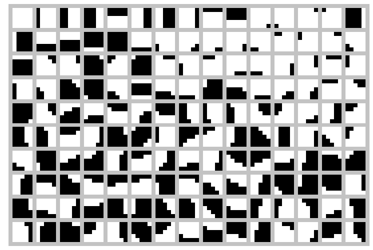


Figure 2: **Clustered patches with character segmentation.**

probability distributions of the intensity of foreground and background pixels respectively.

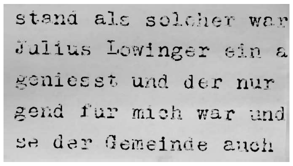
To estimate the probability distributions of the foreground and background pixels, we could use a non-parametric kernel density estimate [7]. However for efficiency reasons though less accurate, we use a parametric estimate, the Gaussian probability function. Differing from the Cao et al. where the Gaussian parameters (mean and variance) are estimated by segmenting y on the fly to obtain a binary image $\hat{x}$, using an adaptive thresholding algorithm, we use an offline estimation process.

Given a test degraded image y for which we seek to estimate its binary image x, we use the gray level histogram of an image $\tilde{y}$ with similar degradations to y whose ground truth image $\tilde{x}$ is known, to obtain more accurate parameters for the foreground and background distributions of y offline. Using the interactive document enhancement system (mentioned in Sec. 2.3), we can label a subset of documents with similar degradation patterns to our target class of document images. By comparing the gray level histogram of $\tilde{y}$ to $\tilde{x}$, we obtain more reliable estimates of the parameters for the foreground and background distributions of test image y. Given the poor level of degradation, labeling of the document images by an human expert to generate $\{\tilde{y}, \tilde{x}\}$ is highly time consuming (hence the need for an automated system), however a single image pair is sufficient to learn the observational model parameters.

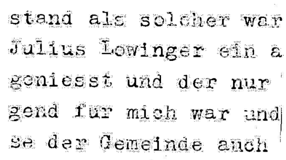
Note that the process of obtaining the Gaussian parameters for the observation model is computed only once and applied to each sub-image during the inference phase of the MRF algorithm. The model does not change for each sub-image.

3. Results and Discussion

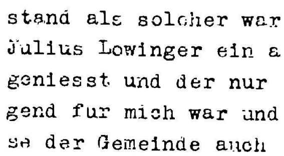
To quantitatively measure the performance of our system in comparison to the other techniques, we tested it on synthetically degraded typewritten ground truth images using Misclassification Error (ME) [8] as the performance criteria. ME is computed as $(M/P) \times 100$, where M is the number of pixels in the enhanced output image $\hat{G}$ that do not correlate with the ground truth



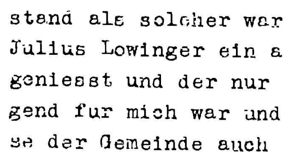
(a) Degraded Image



(b) Heuristic adaptive algorithm

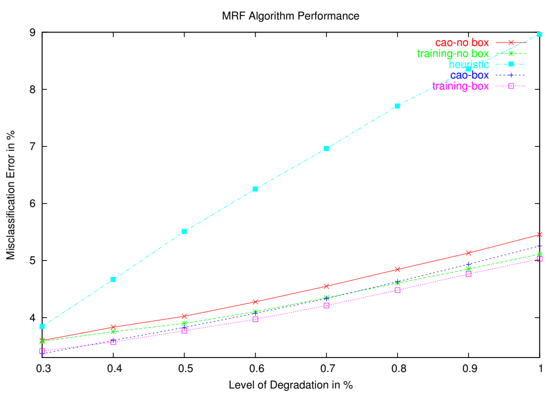


(c) Cao et al. algorithm

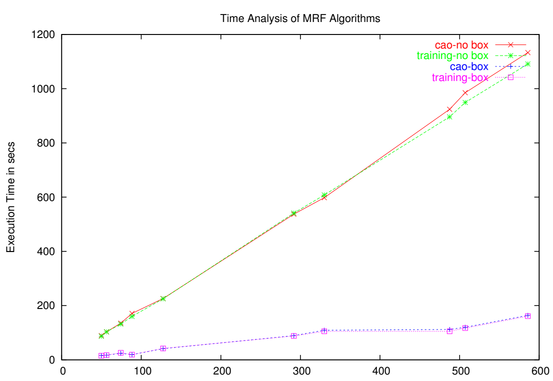


(d) Proposed algorithm

Figure 4: **Visual results of the different algorithms.**



(a) Effect of Character Segmentation on MRF Algorithm



(b) Execution Time Analysis

Figure 3: **Comparison of the different approaches.**

image G and P is the number of pixels in the original binary degraded image D. The degradation model, similar to the one outlined in [3], is modeled after the degradations observed in actual degraded historical typewritten documents. We varied the degree of degradation by varying the percentage of windows selected for random degradation in the image. We evaluate the proposed approach of using character segmentation and training data in observation model to the Cao et al. algorithm (estimation on the fly), with and without character segmentation, and a heuristic adaptive algorithm [4]. Preliminary results of this evaluation on a test dataset of 4 document images are shown in Figs. 3 and. 4.

As can be observed in Fig. 3a, the proposed approach produces a smaller misclassification error compared with known techniques. Time performance of the evaluated approaches is shown in Fig. 3b. As shown, the proposed approach is substantially faster by roughly an order of magnitude. This is due to the use of character bounding boxes which decouples models and greatly reduce the time complexity of the MRF technique. Qualitative results when the degree of degradation is 1.0% for one of test images are shown in Fig. 4.

4. Conclusion

We propose a novel MRF approach for document image enhancement. The proposed approach uses image data for obtaining correct patch alignment and utilizes character segmentation for decoupling generated models. Experimental evaluation using historical documents from the Frieder collection shows that the proposed approach is an order of magnitude faster compared with known techniques while achieving higher effectiveness.

References

[1] The diaries of Rabbi Dr. Avraham Abba Frieder. http://ir.iit.edu/collections/.

[2] A. Antonacopoulos and C. Castilla. Flexible text recovery from degraded typewritten historical documents. In *Proc. ICPR*, pages 1062–1065, 2006.

[3] G. Bal, G. Agam, G. Frieder, and O. Frieder. Interactive degraded document enhancement and ground truth generation. In B. Yanikoglu and K. Berkner, editors, *Document Recognition and Retrieval XV*, volume 6815 of *Proc. SPIE*, 2008.

[4] J. Bernsen. Dynamic thresholding of gray-level images. In *Proc. ICPR*, pages 1251–1255, 1986.

[5] H. Cao and V. Govindaraju. Handwritten carbon form preprocessing based on markov random field. In *Proc. CVPR*, pages 1–7, 2007.

[6] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael. Learning low-level vision. *International Journal of Computer Vision*, 40(1):25–47, 2000.

[7] M. D. Gupta, S. Rajaram, N. Petrovic, and T. S. Huang. Restoration and recognition in a loop. In *Proc. CVPR*, 2005.

[8] M. Sezgin and B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *J. Electron. Imaging*, 2004.

[9] C. Wolf and D. Doermann. Binarization of low quality text using a markov random field model. In *Proc. ICPR*, 2002.

[10] J. Zeng and Z.-Q. Liu. Stroke segmentation of chinese characters using markov random fields. In *Proc. ICPR*, pages 868–871, 2006.