

MULTI-SPECTRAL DOCUMENT IMAGE BINARIZATION USING IMAGE FUSION AND BACKGROUND SUBTRACTION TECHNIQUES

Nikolaos Mitianoudis and Nikolaos Papamarkos

Image Processing and Multimedia Laboratory,
Department of Electrical and Computer Engineering,
Democritus University of Thrace,
67100 Xanthi, Greece

ABSTRACT

In this paper, the authors exploit a multispectral image representation to perform more accurate document image binarisation compared to previous color representations. In the first stage, image fusion is employed to create a “document” and a “background” image. In the second stage, the FastICA algorithm is used to perform background subtraction. In the third stage, a spatial kernel K-harmonic means classifier binarizes the FastICA output. The proposed system outperforms previous efforts on document image binarization.

Index Terms— Document Image Binarization, Image Fusion, Multi-spectral imaging, Independent Component Analysis

1. INTRODUCTION

Document images may arise from a historical document or a book that is digitized using a scanning device or a camera. In order to exploit these document images more efficiently, it is essential to extract their textual information, i.e. perform Optical Character Recognition (OCR). Nevertheless, scanned documents often contain undesired noise due to the document’s wear or the scanning procedure. To enhance the performance of OCR algorithms, these noise elements need to be removed before document image binarization [1]. *Image Binarization* is a process, where the image pixels are classified into into a binary representation (text and background pixels) that will be processed by a character recognition algorithm.

There exist many binarisation approaches. Conventional binarization techniques of gray-scale documents were initially based on global thresholding algorithms, which have proved to be efficient for simple gray-scale images, but are inappropriate for complex and degraded documents. For this purpose, local binarization techniques have been extensively used [2]. One of the most efficient methods is proposed by Gatos et al. (GPP) [3], where the document background is estimated by an adaptive threshold, which labels each pixel as either text or background. To estimate the background surface, they used Sauvola’s binarization algorithm to

approximately extract the text pixels and calculated the background surface by interpolation of neighboring background pixels intensities. In [4], Lu et al. estimate a document background surface through an iterative polynomial smoothing procedure. Then, the document text is segmented by a local threshold, estimated based on the detected text stroke L1-image edges. In [5], Makridis and Papamarkos introduced a two-stage image binarization approach, using a background removal technique that was based on fixed-size median filtering of the document image and a 2D clustering on a novel 2D image representation of neighboring pixels of similar intensity. Recently, the authors extended this work using a Local Co-occurrence Mapping and local contrast information and a two-state Gaussian Mixture Model classifier, outperforming the GPP method [2].

Recently, Hedjam and Cheriet [6] demonstrated a novel methodology of enhancing document image processing using multi-spectral imaging. Their system employs a set of band-pass filters in order to capture high-resolution images of different spectra. The multispectral imaging output corresponds to the Ultra-Violet (UV), visible (VIS) and Infra-Red (IR) wavelengths. Hedjam and Cheriet [6] have acquired over 110 images, containing 17th to 20th century historical documents. Most of the acquired image contain handwritten text with iron-gall base ink, with high degradations. This ink type is known to appear in the color (visible) wavelengths, but to disappear gradually in the IR ones. In contrast, most common degradations continue to appear in the IR wavelengths. Hedjam and Cheriet introduced a new restoration model in which degradation areas are identified by exploiting the information in the infrared channels. A binarization transform is then used to generate a binary mask to locate the areas which should be inpainted. Small degradations are first eliminated, followed by a mechanism to tackle degradations overlapping with text. It is essential to point out that their approach improves the output of other binarization algorithms and does not act as an independent binarization algorithm.

In the proposed system, we perform image binarization using a multi-stage approach. In the first stage, we employ

image fusion to create two images: a) the document image, b) the background image. In the second stage, we use a background subtraction technique, based on Independent Component Analysis (ICA) to remove information belonging to the background image from the document image. This removes many artifacts, including page foldings, tears and stains. In the third stage, we use a spatial K-harmonic means algorithm to perform binarization, taking into account neighborhood information in the clustering procedure. In the final post-processing stage, small-size 8-connected clusters are removed to eliminate possible binarization noise.

2. PROPOSED METHODOLOGY

In this section, we describe each stage of the proposed system.

2.1. Image Fusion stage

In the first stage, we exploit the multi-spectral representation of the document image to create two synthetic images: a) the document image, b) the background image. The initial multi-spectral image dataset, kindly provided online by [6], contains eight bands: one in the ultra-violet (UV) wavelength (340 nm), three bands in the visible wavelength (RED (500 nm), GREEN (600nm), BLUE (700nm)) and four in the Infrared (IR) wavelengths (800nm, 900nm, 1000nm, 1100nm). We observed that the first six wavelengths display text information, whereas the last two contain mainly background information and barely any text. Thus, it is sensible to create two images, one containing all document information and one containing background information. The process of combining the important features from a set of input images to form a single enhanced image is referred to as *Image Fusion*. Transform-domain image fusion approaches tend to outperform spatial-domain ones. In this paper, we employed Mitianoudis and Stathaki's image fusion, which operates on a transform domain, defined by self-trained ICA bases [7]. This image fusion system not only transfers salient information (detail, edges) to the fused image, but can also handle possible different brightness input image levels, by adjusting to an optimal contrast level for the fused image [8].

In essence, we create the "document" image, by fusing the first six wavelengths using the Mitianoudis-Stathaki ICA fusion system with the "max-abs" fusion rule and the optimal contrast optimisation enabled. The "background" image is created by fusing the last two wavelengths using the same algorithm and configuration. To match the intensity levels of the two fused images, we linearly transform the intensities of the "background" image, so that its dominant histogram peak is in the same intensity level as the dominant histogram peak of the "document" image. This linear transformation is chosen instead of non-linear histogram modifications, since we are only interested in matching the average intensities of the two images (background image and the background present in

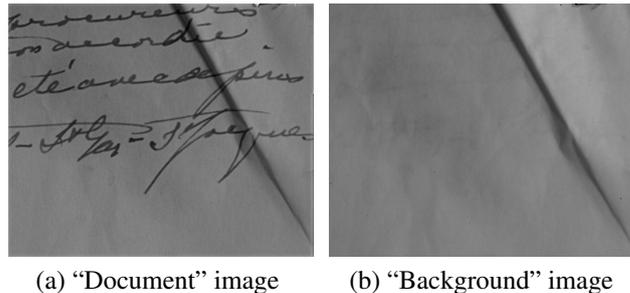


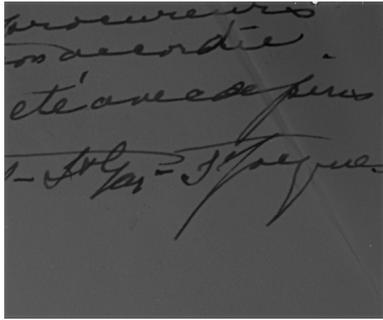
Fig. 1. The "Document" and the "Background" image after the fusion process.

the document image). In Fig. 1, a typical image set created by the fusion stage is depicted. The "background" image serves as a very good estimate of the document's noisy elements that need to be removed.

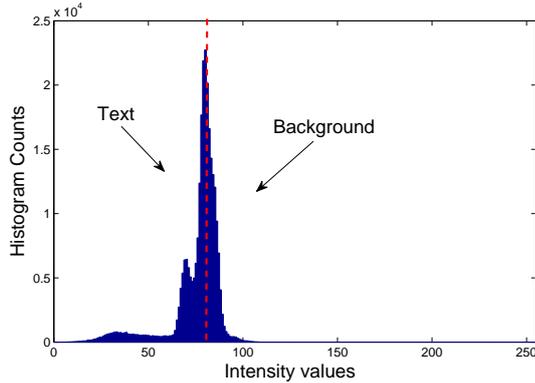
2.2. Background Subtraction stage

In the next stage, we attempt to remove elements of the "background" image from the "document" image. This problem resembles the problem of background subtraction for moving objects tracking in surveillance videos. In these applications, one can obtain a background image estimate from a frame containing no moving objects. This background image is then subtracted from frames containing moving objects in order to separate these moving entities. In [9], Tsai and Lai treat the whole background subtraction as a linear instantaneous blind source separation problem [10], where we need to identify two independent components: the background component and the moving object component. Here, we can use exactly the same setup, since the document image can be considered a linear superposition of the background image and an image containing characters. Thus, we separate the two independent components, background and text, from the two mixtures i.e. the two fusion outputs of the previous stage.

In [9], Tsai and Lai claim that they cannot use traditional ICA algorithms, such as the FastICA algorithm [10], because the probability distribution (histogram) of each component is very different to the ones encountered in speech separation and thus nonlinearities, such as the $\tanh(\cdot)$ that denotes a super-Gaussian profile, cannot be used. Therefore, they attempted to perform ICA by minimising of mutual information of the actual sources, using a particle swarm optimization. Here, to alleviate this problem, we enhance the super-Gaussian characteristics of the signals, by simply moving the input images to the DCT domain, where sparsity is enhanced. Instead of performing DCT for the whole image, we can increase sparsity even more by performing 8×8 block-based DCT (similar to the one used in JPEG) to the input images. The assumed linear instantaneous mixing model is not altered by linear transformations and thus the separation is equiva-



(a) Separated text Component



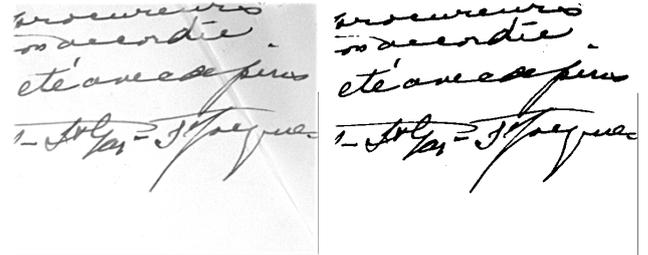
(b) Histogram

Fig. 2. The separated text component and its histogram after the FastICA algorithm.

lent in either domain. The sparse images in the DCT domain tend to have more “heavy-tailed” superGaussian histograms and thus the FastICA algorithm can be properly applied. After separation, the two images are moved again to the spatial domain. In order to choose the text component from the separated ones, we measure the *kurtosis* of the separated components. Since the background component does not contain strong salient information, but rather more noise, i.e. a more Gaussian structure, is thus bound to feature smaller kurtosis values compared to the text component. In Fig. 2 (a), the output of the background subtraction stage is presented.

2.3. Binarization stage

We can see that after the background subtraction stage, there still exist certain artifacts in the image. Before we binarise the image, we observed that some points can still be attributed to background before binarization. We examined the histogram of the separated text component after the FastICA algorithm. In Fig. 2(b), a typical histogram is depicted. We observed that these histograms tend to be skewed to one side of the mean. The pixels that constitute the skewed side is essentially text information, where as the other cluster is background noise



(a) Preprocessing

(b) Binarization Result

Fig. 3. (a) The image after removing negative skewed points, (b) the SKKHM binarization result.

and thus can be set to white. In summary, we measure the skewness of the separated component. The sign of skewness can indicate which side to the mean can be set to white (background). In Fig. 3 (a), we can see the result of this thresholding to the separated text component.

After this step, we can use a clustering technique to binarize the text component. Here, we use a spatial kernel K-harmonic Means clustering (SKKHM) algorithm by Li et al. [11]. This is a modified K-means classifier that employs harmonic distances to overcome the centre initialisation problem. It also uses a kernel-based distance, to make the classifier less sensitive to outliers and noise. Finally, spatial image information, derived from Markov Random Field modelling, is also incorporated to make the clusters more spatially consistent. There are several parameters that control the neighbourhood size Q , the neighbourhood influence magnitude β to cluster formation and the σ of the kernel Radial Basis function (RBF) (for more details, please see [11]). In our experiments, we used a 3×3 neighbourhood, and a value of 0.5 for both β and σ . Finally, isolated blobs or small misclassified noisy items can be removed by discarding small (less than 20 pixels) connected objects with 8-connectivity in the binary output. The final binarization result is depicted in Fig. 3(b).

3. EXPERIMENTS

In this section, we quantify the performance of the proposed binarization system. We used the dataset kindly offered online by Hedjam and Cheriet¹, consisting of 10 multispectral images of handwritten documents. We used the Mitianoudis and Stathaki’s Image Fusion system with 60 trained 8×8 bases on natural images. The “max-abs” rule was used with optimal contrast estimation. For the SKKHM algorithm, we used a 3×3 neighborhood, $\beta = 0.5$ and $\sigma = 0.5$. Both these algorithms are available online². We used the Gatos et al (GPP) [3] binarization implementation, kindly offered by the authors and our recent LCM binarization method [2] for

¹<http://www.synchromedia.ca/databases/HISTODOC1>.

²<http://utopia.duth.gr/~nmitiano>

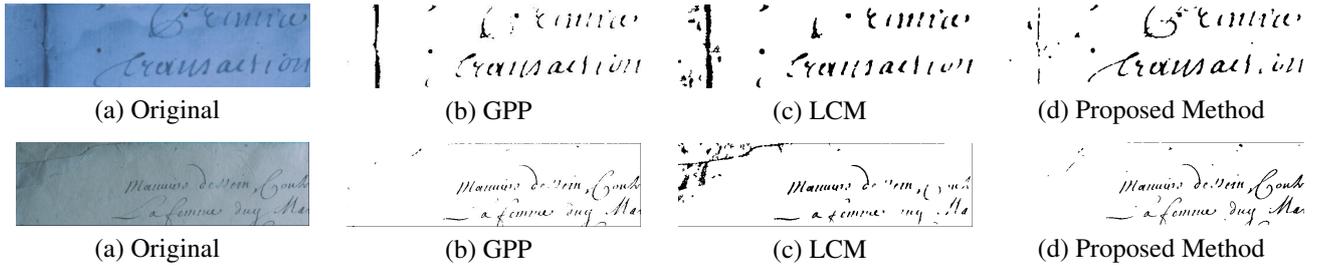


Fig. 4. Some typical binarisation results for the three methods.

Table 1. Precision - Recall and F-Measure results for the GPP, LCM and the proposed method.

Image	GPP			LCM			Proposed		
	Precision	Recall	FM	Precision	Recall	FM	Precision	Recall	FM
SM20	0.84	0.46	0.60	0.71	0.63	0.67	0.86	0.77	0.82
SM30	0.79	0.69	0.74	0.71	0.98	0.82	0.89	0.94	0.92
SM36	0.91	0.59	0.72	0.73	0.94	0.82	0.97	0.74	0.82
SM58	0.83	0.55	0.66	0.63	0.88	0.73	0.77	0.76	0.77
SM60	0.85	0.4	0.54	0.57	0.88	0.69	0.75	0.66	0.70
SM67	0.39	0.45	0.41	0.3	0.53	0.39	0.64	0.67	0.65
SM68	0.92	0.67	0.78	0.39	0.97	0.56	0.82	0.83	0.82
SM76	0.95	0.57	0.72	0.88	0.81	0.84	0.82	0.89	0.85
SM95	0.73	0.38	0.5	0.6	0.66	0.63	0.9	0.7	0.79
Average	0.79	0.53	0.63	0.61	0.81	0.68	0.82	0.77	0.79

performance comparison. As inputs to these algorithms, we use the visible wavelength (RGB) images. Numerical evaluation is conducted using the hand-labeled binary ground truths, provided with the dataset. To quantify performance, we calculated *Precision*, *Recall* and the *F-Measure*, as commonly used in similar benchmarks and defined in [3]. Table 1 outlines the algorithms' average performance as well as individually for each image. In Fig. 4, some representative results are depicted. Binarisation results for all images can be viewed online here³.

We observe that the proposed system outperforms previous single-channel binarization systems, in terms of F-Measure. While GPP scores an average of 63% and LCM one of 68%, the proposed system scores 79%. This is mainly due to the extra multi-channel information and the more accurate background information, the system is able to extract, due to the extra IR channels. In the other approaches, (GPP and LCM), the background is estimated from the single-channel or color image, which often results to inferior background estimates. The proposed system offers a significant performance improvement. In addition, in most images the proposed system offers balanced precision-recall performance, which is usually preferable in binarization systems. Finally, the authors believe that the background subtraction part of

the system can still be improved, since we still observe small background artifacts in the separated text component.

We did not compare our method directly with the results of Hedjam and Cheriet [6], since their method is offered as a preprocessing step to other binarization algorithms, including GPP and possibly our approach. In addition, our approach only attempts to perform binarization, neither inpainting or restoration. Nevertheless, our performance in Table 1 compares favorably with the scores reported in [6] (max reported score FM=0.9235 for a specific image and method).

4. CONCLUSIONS

In this paper, the authors present a complete system to perform document image binarization using multi-spectral document images. Image fusion is used to fuse different wavelength ranges to create a document and a background image. The details of the background image are then separated for the document image using FastICA. The separated text component is binarised using a Spatial Kernel K-Harmonic means algorithm to impose neighbourhood constraints during clustering. In the future, we would like to improve the separation performance of FastICA, perhaps using a sparser representation or a more efficient separation algorithm, in order to improve the algorithm's performance.

³<http://utopia.duth.gr/~nmitiano/icip14.rar>

5. REFERENCES

- [1] N. Papamarkos, "A neuro-fuzzy technique for document binarisation," *Neural Comput. Appl.*, vol. 12, no. 3 - 4, pp. 190–199, 2003.
- [2] N. Mitianoudis and N. Papamarkos, "Local co-occurrence and contrast mapping for document image binarization," in *Proc. 14th Int. Conf. on Frontiers in Handwritten Recognition (ICFHR14)*, Crete Island, Greece, 2014.
- [3] B. Gatos, I. Pratikakis, and S.J. Perantonis, "Adaptive degraded document image binarization," *Pattern Recognition*, vol. 39, pp. 317–327, 2006.
- [4] S. Lu, B. Su, and C.L. Tan, "Document image binarization using background estimation and stroke edges," *Int. Jour. Document Analysis and Recognition*, vol. 13, pp. 303–314, 2010.
- [5] M. Makridis and N. Papamarkos, "An adaptive layer-based local binarization technique for degraded documents," *Int. Jour. Pattern Recognition and Artificial Intelligence*, vol. 24, no. 2, pp. 1–35, 2010.
- [6] R. Hedjam and M. Cheriet, "Historical document imagerestoration using multi-spectral imaging system," *Pattern Recognition*, vol. 46, pp. 2297–2312, 2013.
- [7] N. Mitianoudis and T. Stathaki, "Pixel-based and region-based image fusion schemes using ICA bases," *Elsevier Information Fusion*, vol. 8, no. 2, pp. 131–142, 2007.
- [8] N. Mitianoudis and T. Stathaki, "Optimal Contrast Correction for ICA-based fusion of Multimodal Images," *IEEE Sensors Journal*, vol. 8, no. 12, pp. 2016 – 2026, 2008.
- [9] D.M. Tsai and S.C. Lai, "Independent component analysis-based background subtraction for indoor surveillance," *IEEE Trans. on Image Processing*, vol. 18, no. 1, pp. 158–167, 2009.
- [10] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley, New York, 2001, 481+xxii pages.
- [11] Q. Li, N. Mitianoudis, and T. Stathaki, "Spatial kernel k-harmonic means clustering for multi-spectral image thresholding," *IET Image Processing*, vol. 1, no. 2, pp. 156–167, 2007.