

Luis Rueda
Domingo Mery
Josef Kittler (Eds.)

Progress in Pattern Recognition, Image Analysis and Applications

12th Iberoamerican Congress
on Pattern Recognition, CIARP 2007
Viña del Mar-Valparaiso, Chile, November 2007, Proceedings

LNCS 4756



 Springer

Restoration of Double-Sided Ancient Music Documents with Bleed-Through

Pedro Castro, R.J. Almeida, and J.R. Caldas Pinto

IDMEC/IST, Technical University of Lisbon
 Av. Rovisco Pais, 1049-001 Lisboa, Portugal
 pedro.melo.castro@ist.utl.pt,
 rjalmeida@ist.utl.pt,
 jcpinto@dem.ist.utl.pt

Abstract. Access to collections of cultural heritage is increasingly becoming a topic of interest for institutions like libraries. With the easy access to information provided by technologies such as the Internet, new ways exist for consulting ancient documents without exposing them to more dangers of degradation. One of those types of documents is written ancient music. These documents suffer from multiple kinds of degradation, where bleed-through outstands as the most damaging. This paper proposes a new method based on the Takagi Sugeno fuzzy classification algorithm to classify the pixels as bleed-through, after performing a general background restoration. This method is applied to a set of double-sided ancient music documents, and the obtained results compared with methods present in the literature.

Keywords: Ancient Music Restoration, Image Processing, Document Degradation, Bleed-through Removal, Registration, Adaptive Thresholding, Fuzzy Classification, Clustering.

1 Introduction

One of the main current goals for institutions like libraries is to give access to its rich collections of cultural heritage, written documents in particular. However, much of those documents suffer from different kinds of degradations that are not corrected in the digitization process. As such, there is an active research on automatic methods to restore these kinds of documents.

Restoration can be seen as a transformation that gives the original aspect to documents showing a certain state of degradation. Degradation, on the other hand, can be described as “every sort of less-than-ideal properties of real document images, e.g. coarsening due to low digitizing resolution, ink/toner drop-outs and smears, thinning and thickening, geometric deformations, etc” [1]. Restoration is necessary not only to enhance the visual appearance of a document, but also to improve the results of further segmentation and recognition operations.

Degradation can be divided into three types [2], according to the parts of a document that are subject to interference: 1) background degradation;

2) foreground degradation; and 3) global degradation. We are concerned with background degradation, which accounts for the majority of defects found in documents of ancient music. It includes blotches due to humidity, marks resulting from ink that traverses the paper (bleed-through) or resulting from the scanning process (show-through), underlines, strokes of pen, annotations, and the superimposition of other symbols. Examples of these degradations are depicted in Fig. 1.

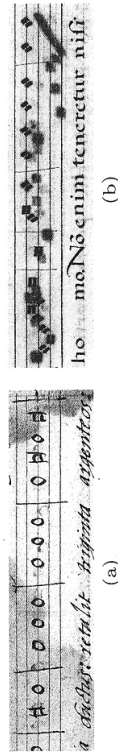


Fig. 1. Images of ancient music showing background degradation

Current approaches for bleed-through removal are split between blind and non-blind methods. Blind methods process a single side of a leaf of paper – the recto – while non-blind methods use both sides – the recto and the verso. Blind methods have the immediate advantage of not requiring the verso image to be present. Non-blind methods, however, dispose of additional features that can be extracted by processing both sides simultaneously, which can lead to a better segmentation.

A previous work by Castro and Pinto [3] has analyzed and compared different recto-based approaches to restore images of ancient music, including global and adaptive thresholding, color clustering, and edge detection. One of the work’s conclusions was that none of the methods was able to successfully remove the defect that resulted from sipped ink, i.e., bleed-through. This therefore demands for a specialized treatment, which is the main focus of this paper.

Non-blind methods are typically comprised of: 1) matching of the recto and verso images; 2) heuristics to distinguish bleed-through from the valid foreground; 3) bleed-through removal; 4) inpainting [4], to fill in the removed bleed-through regions; and 5) pre and/or post-processing algorithms. Sharma [5] developed a simple model of the show-through effects, assuming that the distortion in the documents is due to show-through and that the impairment when scanning such documents can be modeled by the properties of the physical scanning process. However, this method does not apply to the case of bleed-through due to its nature in the derivation of a physical model that describes the show-through phenomenon. Wang and Tan [6] proposed a method that matches pixels on the two sides to remove interference and corrects the result using a Canny edge detector [7]. However, this correction is favored by the orientation of the characters, which cannot be applied to the case of written ancient music as these are typically written in a nearly vertical orientation. Dubois and Dano [8] segment a document into four regions: foreground only, bleed-through only, background, and foreground and bleed-through overlap. Bright areas are considered to be background, whereas areas that are dark on the side of interest and not on

the reverse are considered to be foreground. Areas that are dark on both sides, but more so on the reverse side, are considered to be bleed-through, while if they are dark on both sides but similar they are considered regions of overlap. Thresholds are used to determine if one side is sufficiently darker than the other, and correlations are used to determine whether they are similar. They state that their segmentation scheme is empirical and could therefore be improved.

The rest of this paper is organized as follows. Section 2 details our proposed method. Section 3 presents and analyzes the results, comparing them to an existing method. Section 4 concludes the paper and indicates future work directions.

2 Method

The method we propose processes the images corresponding to the two sides of a leaf of paper in order to restore them. General deterioration, like water blotches and inhomogeneous backgrounds, is removed, but the main focus is on eliminating bleed-through.

The images are converted to gray scale before processing. Besides this conversion, the verso image is flipped horizontally in order for its coordinate system to match that of the recto.

The method is comprised of multiple steps. First, the recto and the verso images are registered. Background homogenization is then performed using adaptive thresholding. Following thresholding, the staff lines are detected. Finally, features are extracted and classification is performed, after which the method proceeds with a final post-processing step.

The result of the method is a binary image that distinguishes from the foreground and the background. All the detected deterioration, including bleed-through, is incorporated into the background. Throughout the exposition of the method, the recto image will stand as the target for restoration, with the verso image being used, in addition, to restore it. The restoration of the verso image follows the exact same approach and therefore will not be covered here.

Problem Formulation. We formulate the problem using a notation that is similar, in part, to that found in the work of Dubois and Dano [8]. Let $f_r^o(x, y)$ and $f_v^o(x, y)$ denote the original recto and verso digital images, respectively, after being converted to gray scale and flipping the verso. The points (x, y) lie on a two-dimensional rectangular space and the range of the functions is the interval $[0, 1]$, with 0 and 1 corresponding to white and black, respectively. Each of the two images is comprised of foreground, background and bleed-through areas. The foreground area contains the writing that was intentionally applied to the paper. The bleed-through area contains part of the verso's writing that has bled through the page. The remainder of the recto corresponds to the background. The original recto image can therefore be represented as

$$f_r^o(x, y) = f_{rfg}^o(x, y) + f_{rbt}^o(x, y) + f_{rbg}^o(x, y), \quad (1)$$

where $f_{rfg}^o(x, y)$, $f_{rbt}^o(x, y)$ and $f_{rbg}^o(x, y)$ correspond to the original recto foreground, bleed-through and background areas, respectively.

This allows us to model the relationship between the two sides:

$$f_r^o(x, y) = C(f_{rfg}^o(x, y), f_{rbg}^o(x, y), f_{vfg}^o(x, y)), \quad (2)$$

where $f_{vfg}^o(x, y)$ denotes the original verso foreground and C is a function that combines its arguments in some way. Possible models for this combining function were presented by Dubois and Pathak [9]. A simple additive model defines it as

$$C(\mu, \rho, \tau) = \mu + \rho + \alpha \cdot \tau, \quad (3)$$

where α represents the attenuation of the verso foreground sipping to the recto.

Still, our method does not highly depend on a specific model of the bleed-through effect. Its ideal purpose is to create a restored recto image, which we denote as $f_r(x, y)$, that nullifies both the background and the bleed-through areas, maintaining only the valid writing:

$$f_r(x, y) = C(f_{rfg}^o(x, y), 0, 0) = f_{rfg}^o(x, y). \quad (4)$$

Conceptually, we wish to determine the parts of an image that correspond to the valid intentional writing. To this accomplishment, the background and bleed-through areas need to be removed. As the bleed-through areas depend on the verso's writing, information from the verso needs to be incorporated and combined with the recto. In practice, however, it is not possible to establish a perfect relation between the two sides. On one hand, the sipping of ink is irregular, as only part of the verso writing does sip to the recto and the bleed-through area may be somewhat diffuse, thus occupying a larger area in the recto than it originally occupied in the verso. On the other hand, ink may have sipped to areas of the recto that already contained valid writing, thereby making it extremely difficult, even for human readers, to distinguish the foreground from bleed-through. Therefore, the main purpose of our method is to restore as much deterioration as possible, recognizing the problem's difficulties beforehand.

Registration. The necessity for using registration comes from the variability in the way the recto and the verso images are captured. When scanning the recto and the verso using typical machines, differences may arise in the resulting images, when compared to each other. This includes mostly shifting, rotation and some skewing. These properties can be modeled by an affine transformation A_t , of parameter vector $t = [t_{11} \ t_{12} \ t_{13} \ t_{21} \ t_{22} \ t_{23}]$ [10], defined as [8]

$$(A_t f)(x, y) = f(t_{11}x + t_{12}y + t_{13}, t_{21}x + t_{22}y + t_{23}). \quad (5)$$

The parameter vector is estimated by solving the optimization problem:

$$\hat{t} = \arg \min_t \sum_x \sum_y \left[f_r^o(x, y) - (A_t f_v^o)(x, y) \right]^2. \quad (6)$$

The registered verso image, i.e., the image that results after applying the affine transformation to the original verso, can then be calculated with

$$f_v^R(x, y) = (A_{\hat{t}} f_v^o)(x, y), \quad (7)$$

while the recto image remains unchanged, thereby having its registered version represented as

$$f_r^R(x, y) = f_r^O(x, y) \tag{8}$$

Background Homogenization. After registering the images, the registered recto background f_r^{rbg} is estimated and removed, along with the degradation it contains. Sauvola's adaptive thresholding method [11] was chosen, as it not only aims at dealing with cases in which the background contains light texture, big intensity variations and uneven illumination, but also has been shown to perform well with written ancient music [3]. This algorithm calculates a threshold value for each pixel based on the mean and standard deviation of all the pixels in a local neighborhood. A window of size $W \times W$ is moved over the image and the threshold value, for a pixel (x, y) , is calculated as

$$t(x, y) = m(x, y) \cdot \left[1 + K \cdot \left(\frac{s(x, y)}{R} - 1 \right) \right], \tag{9}$$

where $m(x, y)$ and $s(x, y)$ are the mean and standard deviation values, respectively, in a local neighborhood of size $W \times W$ of pixel (x, y) , and R is the dynamic range of standard deviation. The values of $W = 15$, $K = 0.2$ and $R = 128$ were used, as they provided the best results in previous experiments [3].

We denote f_r^B as the binary image that results after applying Sauvola's thresholding to f_r^R . This image contains all background pixels set to 0 and the remaining pixels set to 1. It will be the task of the classification step to distinguish from the registered recto foreground and bleed-through, now that the background has been detected.

Staff Line Detection. Staff lines are detected because line pixels may become incorrectly classified as bleed-through in the classification step, leading to broken lines. This allows to restore those lines as a post-processing step.

We used a method that has been shown to perform well with written ancient music [12]. This method uses horizontal projections and small rotations of f_r^B , finding peak areas of the projections and classifying them as staff lines.

Features. The extraction of features was accomplished by taking both the recto f_r^R and the verso f_v^R into account. All the pixel positions (x, y) were selected as possible candidates according to the following criteria:

$$\forall_{x,y} \left[((f_r^B(x, y) = 1) \wedge (f_r^R(x, y) < f_v^R(x, y))) \rightarrow ((x, y) \in \text{Candidates}) \right]$$

From these candidates, 4 features were extracted: 1) correlation coefficients; 2) differences; 3) recto values; and 4) verso values. The correlation coefficients relate to information about whether the verso is similar to the recto, near a certain candidate pixel. For each candidate position, a window of size 9×9 is centered on the recto and the verso, forming two matrices A and B , respectively. The correlation coefficient for a given candidate is then calculated as

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2)(\sum_m \sum_n (B_{mn} - \bar{B})^2)}} \tag{10}$$

where A_{mn} and B_{mn} are pixels with coordinates (m, n) within the windows A and B , respectively, and \bar{A} and \bar{B} are the means of the pixels within A and B , respectively.

The differences correspond to the distances between the pixel values on the two sides. They relate to the subtraction of the recto from the verso, i.e., $f_v^R - f_r^R$. The last two features correspond to the pixel values in the recto and the verso.

Classification. In a classification problem the aim is to learn the behavior between the input and output of the training data. The use of fuzzy models in classification problems has been adopted in many domains [13,14,15], because they are able to solve difficult problems, exhibit robust behavior and present linguistic representations, which are easy to interpret.

The fuzzy models used in this paper are an extension of the Takagi-Sugeno fuzzy models [16] in the affine form. This fuzzy classification rule is a fuzzy if-then rule whose consequent part is a class label [14,15]. It can be described by

$$R^k : \text{If } x \text{ is } A^k \text{ then } x \in \text{class}_k \text{ with confident value } CV^k, \tag{11}$$

where $k = 1, 2, \dots, K$, $i = 1, \dots, K$, K denotes the number of rules in the rule base, R_i is the i th rule, n is the number of features, A_{i1}, \dots, A_{in} are fuzzy sets defined in the antecedent space, y_i is the output feature for rule i , a_i is a parameter vector and b_i is a scalar offset, κ is the number of classes, and CV^k is the confident value of the rule R^k . The confident value of the if-then rule represents the rule weight interpreted as its confident strength. This type of model is used because it focuses on the precision of the obtained model.

To form the fuzzy system model from the data set with N data samples, given by $X = [x_1, x_2, \dots, x_N]^T$, $Y = [y_1, y_2, \dots, y_N]^T$ where each data sample has a dimension of n ($N >> n$), first the structure is determined and afterwards the parameters of the structure are identified. The number of rules characterizes the structure of a fuzzy system. Fuzzy clustering in the Cartesian product-space $X \times Y$ is applied to partition the training data. The partitions correspond to the characteristic regions where the system's behavior is approximated by local linear models in the multidimensional space. Given the training data X_T and the number of clusters K , a suitable clustering algorithm is applied. In this paper we used the fuzzy c-means [17], one of the most widely used clustering algorithms.

As result of the clustering process a fuzzy partition matrix $U = [\mu_{ik}]$ is obtained. The fuzzy sets in the antecedent of the rules are identified by means of the partition matrix U which has dimensions $[N \times K]$. One-dimensional fuzzy sets A_{ij} are obtained from the multidimensional fuzzy sets by projections onto the space of the input variables x_j . This is expressed by the point-wise projection operator of the form $\mu_{A_{ij}}(x_{jk}) = \text{proj}_j(\mu_{ik})$. The point-wise defined fuzzy sets A_{ij} are then approximated by appropriate parametric functions. The consequent parameters for each rule are obtained by means of linear least square estimation, which concludes the identification of the classification system.

Post-processing. Having classified all pixels as bleed-through or foreground, we now have f_r^{fbg} and f_r^R , where the first corresponds to the registered image

without bleed-through. As a final step, two operations need to be performed. First, the initially detected staff lines are restored. Second, bleed-through diffusion, visible as a set of pixels surrounding bleed-through, is suppressed. To this intent, a window of size 5×5 is centered on each bleed-through pixel and all the pixels connected to it, inside that window, showing intensity levels at most 0.04 darker or lighter than the center pixel value, are marked as bleed-through.

We now have the final restored image f_r , containing the detected valid writing, which is derived from the detected foreground f_{rfg}^r , according to (4) and (8), as

$$f_r(x, y) = f_{rfg}^o(x, y) = f_{rfg}^r(x, y). \quad (12)$$

3 Results

Experiments were conducted in order to determine whether our method achieves a good segmentation. The method of Dubois and Dano [8] was also tested, and the results of the two methods compared. The quality of the restoration, as observed in the resultant images, was the main emphasis of our tests. Computational issues were therefore not considered.

Methodology. Images of written ancient music were provided by the Portuguese National Library and the Biblioteca Geral da Universidade de Coimbra, Portugal. A total of 14 images, i.e., 7 pairs of recto-verso images, scanned at a resolution of 150 dpi, were used throughout the experiments. These images contain diverse degradation types, as well as different musical notations and illumination characteristics. They are representative of the majority of images present in the studied collections from the two referred libraries. All images were first manually restored, using graphics editing software, in order to be used as a standpoint for comparison. From the 14 images, one side of each pair was used for training and the other for validation.

It must be noted that the method of Dubois and Dano [8] does not perform binarization. Its purpose is to detect and remove bleed-through areas, preserving the remaining parts of the image. Therefore, for it to be comparable to our method, Sauvola's thresholding was applied to it. The image is thresholded after registration, but the algorithm proceeds as normal, ignoring the thresholded image. Only in the end, when bleed-through pixels have been detected, thresholds are combined to form the resulting image, which constitutes of the thresholded image with the detected bleed-through pixels removed.

The test images were processed by the chosen methods and compared to the manually restored images. The comparison was evaluated by the standard measures of *precision* and *recall* [18], with a slight modification. To evaluate text segmentation, these measures are typically used with the precision of a character or word. In the context of ancient music, this does not apply so well as the musical notation is varied, including notes, clefs, key and time signatures, rests, bar and staff lines, as well as text, among other symbols. All of these symbols could be treated as characters, but that would ignore the great differences in shapes and

sizes among them. Therefore, we opted to perform a bitwise comparison and, as such, the *precision* (P) and *recall* (R) measures were used as

$$P = \frac{\text{Correctly Detected BT}}{\text{Total Detected BT}} \quad \text{and} \quad R = \frac{\text{Correctly Detected BT}}{\text{Total BT}}, \quad (13)$$

where "Correctly Detected BT" refers to the bleed-through pixels that were correctly binarized (i.e., that are equal to those of the manually restored images), "Total Detected BT" refers to the total bleed-through pixels that were binarized, and "Total BT" refers to the total bleed-through pixels that are present in the manually restored images.

Precision and *recall* reflect the performance of removing interfering strokes and restoring valid strokes, respectively. To relate the two measures, the geometric mean (g -mean) was used, being defined as

$$g\text{-mean} = \sqrt{P * R}. \quad (14)$$

This measure was used because it does not depend on the distribution of examples between classes [19]. This is convenient as the number of bleed-through pixels is typically a minority, when compared to the entire set of pixels within an image.

Results. The results of evaluating our method, which we denote as CAP, and that of Dubois and Dano, denoted as DD, on the 7 selected images, are presented in Table 1. Our method achieved a good overall segmentation, with a better performance in almost all images. It should be noted that there is an inherent degree of error in these results, as they are based on a pixel-wise comparison with manually restored images. When restoring those images by hand, it is hard to determine the exact class for each pixel, as the value of some pixels is not visually distinct.

Table 1. Detailed *precision* (P), *recall* (R) and g -mean (G) results obtained by applying the methods to 7 images of ancient music

Image	1	2	3	4	5	6	7	Average
P	0.249	0.719	0.487	0.713	0.514	0.400	0.025	0.444
R	0.921	0.814	0.718	0.770	0.795	0.833	0.788	0.806
G	0.479	0.765	0.591	0.741	0.639	0.577	0.139	0.562
P	0.135	0.423	0.254	0.667	0.230	0.241	0.032	0.283
R	0.832	0.737	0.509	0.807	0.859	0.845	0.894	0.783
G	0.335	0.558	0.360	0.734	0.444	0.452	0.169	0.436

Parts of two images used throughout the tests are presented in Fig. 2. It can be noted that the readability is greatly improved with the removal of bleed-through areas. It can also be observed that the removal is naturally not perfect, as some bits of diffusion still remain. However, these bits are dispersed, therefore not creating a serious visual impact to the reader.

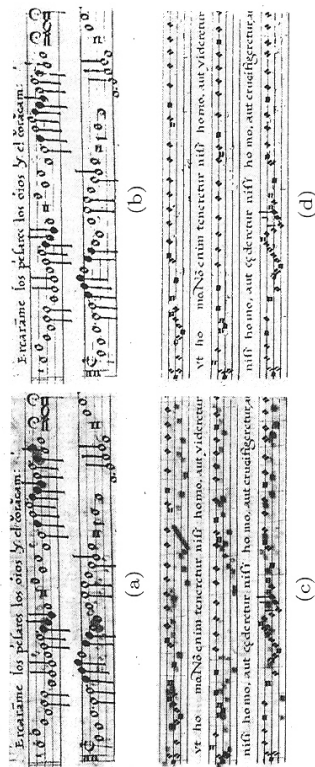


Fig. 2. Results of processing two of the original images with the proposed method. The first column presents the original images, while the second shows the restoration results.

4 Conclusion

Written ancient music images often present multiple types of degradation. Typical cases of degradation were presented and analyzed, including the specialized case of bleed-through. A restoration method for images of ancient music suffering from bleed-through has been proposed. A methodology for method evaluation was established and the method compared to an existing one. The results demonstrate that the proposed method performs well and greatly improves the visual appearance of the documents.

Some work is still necessary to fine-tune the proposed method. This includes the diffusion removal in the post-processing step, as well as a general performance optimization that should be considered. Furthermore, some research is still necessary to fully automate the application of this method to the mass restoration of written ancient music. This will be the concern of future work.

Acknowledgments. This work was partly supported by: “Programa de Financiamento Plurianual de Unidades de I&D (POCTI), do Quadro Comunitário de Apoio III”; the FCT project POSC/EIA/60434/2004;(CLIMA), Ministério do Ensino Superior da Ciência e Tecnologia, Portugal; program FEDER; and “Programa Operacional Ciência e Inovação POCI2010”.

References

1. Baird, H.: The state of the art of document image degradation modeling (2000)
2. Driira, F.: Towards restoring historic documents degraded over time. In: Document Image Analysis for Libraries, pp. 350–357 (2006)
3. Castro, P., Pinto, J.R.C.: Methods for written ancient music restoration. Image Analysis and Recognition. In: Forth International Conference ICIAR (to be published, 2007)

4. Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: SIGGRAPH-2000 pp. 417–424 (2000)
5. Sharma, G.: Show-through cancellation in scans of duplex printed documents. *IEEE Transactions on Image Processing* 10(5), 736–754 (2001)
6. Wang, Q., Tan, C.L.: Matching of double-sided document images to remove interference. In: CVPR, pp. 1084–1089. IEEE Computer Society, Los Alamitos (2001)
7. Canny, J.: A computational approach to edge detection. In: RCV 1987, 184–203 (1987)
8. Dubois, E., Dano, P.: Joint compression and restoration of documents with bleed-through. In: IS&T Archiving, pp. 170–174 (2005)
9. Dubois, E., Pathak, A.: Reduction of bleed-through in scanned manuscript documents. In: PICS, pp. 177–180 (2001)
10. Brown: A survey of image registration techniques. *CSURV: Computing Surveys* 24 (1992)
11. Sauvola, J., Pietikainen, M.: Adaptive document image binarization. *Pattern Recognition* 33(2), 225–236 (2000)
12. Pinto, J.R.C., Vieira, P., da Costa Sousa, J.M.: A new graph-like classification method applied to ancient handwritten musical symbols. *IJDAR* 6(1), 10–22 (2003)
13. Setnes, M., Roubos, H.: GA fuzzy modeling and classification: Complexity and performance. *IEEE Transactions on Fuzzy Systems* 8(5), 509–522 (2000)
14. Hirota, K., Pedrycz, W.: Fuzzy computing for data mining. *Proceedings of the IEEE* 87(9), 1575–1600 (1999)
15. Thawonmas, R., Abe, S.: Function approximation based on fuzzy rules extracted from partitioned numerical data. *IEEE Transactions on Systems, Man and Cybernetics, Part B* 29(4), 525–534 (1999)
16. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its application to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics* 15(1), 116–132 (1985)
17. Bezdek, J.C.: *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York (1981)
18. Junker, M., Dengel, A., Hoch, R.: On the evaluation of document analysis components by recall, precision, and accuracy. In: ICDAR, pp. 713–716 (1999)
19. Kubat, M., Holte, R., Matwin, S.: Machine learning for the detection of oil spills in satellite radar images. *Machine Learning* 30, 195–215 (1998)