ENHANCED BLEEDTHROUGH CORRECTION FOR EARLY MUSIC DOCUMENTS WITH RECTO-VERSO REGISTRATION

John Ashley Burgoyne   Johanna Devaney   Laurent Pugin   Ichiro Fujinaga
Centre for Interdisciplinary Research in Music and Media Technology
Schulich School of Music of McGill University
Montréal (Québec) Canada
{ashley,devaney,laurent,ich}@music.mcgill.ca

ABSTRACT

Ink bleedthrough is common problem in early music documents. Even when such bleedthrough does not pose problems for human perception, it can inhibit the performance of optical music recognition (OMR). One way to reduce the amount of bleedthrough is to take into account what is printed on the reverse of the page. In order to do so, the reverse of the page must be registered to match the front of the page on a pixel-by-pixel basis. This paper describes our approach to registering scanned early music scores as well as our modifications to two robust binarization approaches to take into account bleedthrough and the information available from the registration process. We determined that although the information from registration itself often makes little difference in recognition performance, other modifications to binarization algorithms for correcting bleedthrough can yield dramatic increases in OMR results.

1 MOTIVATION AND BACKGROUND

1.1 Fostering Interdisciplinary Research with OMR

"We stand at a moment of opportunity," opened Nicholas Cook at his invited talk for ISMIR 2005 in London. The opportunity is for historical musicologists and music information scientists to work together and revitalize the subdiscipline of computer-assisted empirical musicology [6]. This subdiscipline began in the 1960s [15], and although it has developed into a thriving discipline in several regions, it is largely moribund in North America, where a significant amount of other musicological research takes place. While Cook assigned musicologists a great deal of the responsibility for realizing this moment of interdisciplinary opportunity, he challenged researchers in music information retrieval to create large databases of the "highly reduced data"—e.g., scores—upon which musicological research relies. Such electronic databases are especially important for those older documents that are available in only a limited number of locations and for which archivists often restrict physical access, making it difficult to engage in large-scale comparative research.

Entering musical sources into such databases by hand is highly labor-intensive, which renders music digitization projects prohibitively costly for most institutions [3]. Optical music recognition (OMR), the musical analog to optical character recognition (OCR), is the most practical means by which to create such databases. The potential for optical recognition to transform research approaches has already been demonstrated by the recent explosion in the number of searchable electronic texts available for older books and journal articles, (e.g., JSTOR [1]). When treating historical documents, however, OMR and OCR systems struggle with various types of document degradation, including ink bleedthrough from the reverse side of the page [1]. Because these systems rely on the ability to distinguish foreground (ink) from background (paper), the darker the bleedthrough, the more likely it is that the bleedthrough will be classified as foreground and thereby corrupt the recognition process.

1.2 Binarization and Bleedthrough

Binarization is the common name for separating an image into its foreground and background. It is notoriously difficult to evaluate, leading most authors to resort to coarse subjective distinctions such as "better," "same," or "worse," e.g., [12]. As binarization is typically a preprocessing step in an automated image-processing pipeline with some other goal, e.g., OMR, one can use the evaluation metric for the ultimate goal, e.g., minimizing the amount of time necessary for a human editor to correct recognition errors, to evaluate the quality of the binarization algorithm. This paper uses a similar method to that described in our recently-published survey of binarization algorithms for music documents [4] to evaluate extensions to the algorithms we found to be best-performing.

Although there has been some previous work on binarization in the presence of bleedthrough that focuses specifically on historical documents [9, 10], a larger body of work

[1] http://www.jstor.org/

exists on the related problem of *showthrough* from the reverse side caused by the bright light used in photoreproduction (e.g., photocopying or scanning). Many approaches to showthrough (and bleedthrough) require that the recto and verso sides of each page be aligned, or *registered*, prior to the removal of showthrough [5, 8, 11, 20]. In addition to "blind" extensions to existing algorithms for the treatment of bleedthrough, i.e., those that do not take into account information from the reverse side of the page, we have developed our own technique for registering the recto and verso and extended the top-performing blind binarization algorithms for music to account for this information.

2 IMAGE PROCESSING

2.1 Binarization

2.1.1 KL Thresholding

In most cases, binarization algorithms seek to identify a global threshold for a given image. All pixels with gray levels below this threshold are classified as foreground and all pixels with gray levels above it are classified as background. The motivation for this simple classification technique, known as *thresholding*, is its computational speed.

Previous studies have demonstrated that thresholding approaches based on Shannon entropy are strong performers [4, 7]. For music documents in particular, [4] shows that a pair of related algorithms, [2] and [13], are optimal. These two algorithms consider images to be probability density functions, the so-called "monkey model," under which gray levels represent the probability that imaginary balls of luminance tossed by monkeys in a uniformly-distributed fashion would land on a particular pixel. Given a threshold T, the binarization of an image is represented not as zeros and ones but as the average gray levels of the foreground and background pixels in the original image: $\mu_0(T)$ and $\mu_1(T)$. Under this representation, the total luminance of the original image and the binarized image is identical. Thus, one can minimize the cross-entropy between the two images, or equivalently, minimize the Kullback-Leibler (KL) divergence, by minimizing

$$\theta(T) = \sum_{f(x,y) \le T} f(x,y) \log \frac{f(x,y)}{\mu_0(T)} + \sum_{f(x,y) > T} f(x,y) \log \frac{f(x,y)}{\mu_1(T)} \quad (1)$$

where $f(x, y)$ represents the gray level of the image at pixel (x, y). This algorithm is the one proposed in [13] and is referred to in this paper as asymmetric KL thresholding. The algorithm proposed in [2] is very similar except that it minimizes the symmetrized form of the KL divergence:

$$\theta(T) = \sum_{f(x,y) \le T} f(x,y) \log \frac{f(x,y)}{\mu_0(T)} + \sum_{f(x,y) \le T} \mu_0(T) \log \frac{\mu_0(T)}{f(x,y)} + \sum_{f(x,y) > T} f(x,y) \log \frac{f(x,y)}{\mu_1(T)} + \sum_{f(x,y) > T} \mu_1(T) \log \frac{\mu_1(T)}{f(x,y)} \quad . \quad (2)$$

We call this technique symmetric KL thresholding. Both algorithms are analogous to Otsu's famous thresholding algorithm, [17], which also seeks to minimize the difference between the original image and a binarized representation of mean gray levels $\mu_0(T)$ and $\mu_1(T)$ but uses mean squared error as opposed to the KL divergence.

In an attempt to account for bleedthrough, which is usually lighter than foreground, we have added a third class to these algorithms by adding a second threshold U. Pixels with gray levels between T and U are considered to be bleedthrough and pixels with gray levels above U are considered to be true background. A new gray-level mean, $\mu_2(T, U)$ is computed for the bleedthrough pixels. For the asymmetric case, we minimize over both thresholds thus:

$$\theta(T,U) = \sum_{f(x,y) \le T} f(x,y) \log \frac{f(x,y)}{\mu_0(T)} + \sum_{T < f(x,y) \le U} f(x,y) \log \frac{f(x,y)}{\mu_2(T,U)} + \sum_{f(x,y) > U} f(x,y) \log \frac{f(x,y)}{\mu_1(U)} \quad . \quad (3)$$

We also implemented the analogous modification to (2). In this paper, these variants are referred to as three-class symmetric and three-class asymmetric KL thresholding. Samples of the output from these algorithms appear in Figure 1.

2.1.2 Gatos et al. 2004

Gatos et al. [10] is an adaptive technique for digital image binarization consisting of three stages: preprocessing, rough foreground estimation, background surface estimation, and final thresholding. In our earlier experiments [4], it was the only locally adaptive algorithm to perform well for music documents. Rough foreground estimation is achieved through the application of the adaptive thresholding method in [16], which tends to produce an over-complete (noisy) estimate of the foreground. A background surface is initially created from pixels classified as background during

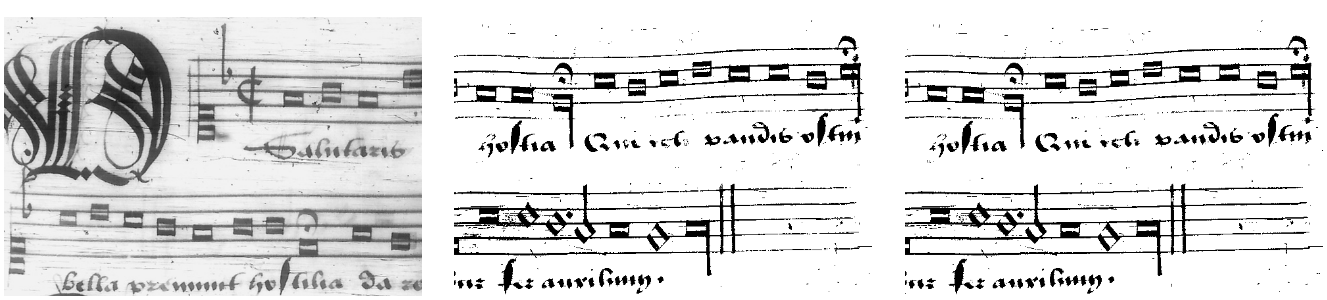
(a) Original recto image (b) Asymmetric KL: 2 classes (c) Symmetric KL: 2 classes

(d) Gatos et al. (21-px window) (e) Gatos et al. (31-px window) (f) Symmetric KL: 3 classes

Figure 1: Binarization output from selected algorithms on a page of the Occo Codex.

(a) Original recto image (b) Original verso image (c) Absolute difference between the original recto image and the registered, flipped verso image

Figure 2: Registration output for RISM M-0539



(a) Original verso image (b) Symmetric KL: recto-verso (c) Gatos et al. (31-px window): recto-verso

Figure 3: Binarization output from selected recto-verso algorithms on the same page of the Occo Codex.

this preliminary thresholding, and the values for the remaining pixels are derived by interpolating between the values of neighboring pixels. The final binarization is achieved by first calculating the distance between the background surface estimate with the original grayscale image: when an original pixel's gray level is greater than a threshold value derived from the corresponding background estimate, it is considered to be part of the foreground. Samples of the output from this algorithm also appear in Figure 1.

2.2 Registration

As the recto and verso side of a page are scanned separately, there is often some degree of misalignment and distortion between them. In order to take advantage of knowledge of the verso when correcting for bleedthrough, it is necessary to register the verso to the recto. We employ a feature-based registration technique, following from [8], that uses an affine transformation to minimize the total squared difference between pixels of recto and verso. In our experiments we constrained the affine transformation to avoid excessive translation or rotation. We also optimized the technique with the simulated-annealing approach described in [11]. Based on the success of the KL-based thresholding algorithms for music images, however, we replaced the sum-of-squares error metric used in [8] and [11] with symmetrized KL divergence.

Incorporating registration information into KL thresholding algorithms is similar to the three-class variant of the algorithms. Again, we minimize over two thresholds, T for the original image and U for the registered version of the reverse side. The bleedthrough class of pixels is restricted to those for which U classifies the pixel on the reverse side as foreground:

$$\begin{aligned}\theta(T,U) = & \sum_{f(x,y)\le T} f(x,y)\log\frac{f(x,y)}{\mu_0(T)} \\ & + \sum_{f(x,y)>T\,\&\,g(x,y)\le U} f(x,y)\log\frac{f(x,y)}{\mu_2(T,U)} \\ & + \sum_{f(x,y)>T\,\&\,g(x,y)>U} f(x,y)\log\frac{f(x,y)}{\mu_1(T,U)} \quad , \qquad (4)\end{aligned}$$

where $g(x,y)$ represents the gray level of the image at the pixel from the registered reverse side corresponding to front-side pixel (x,y). Again, we also implemented the analogous modification to (2) for the symmetric form.

In order to take advantage of the registered verso in Gatos et al., we added an additional stage to the process between the foreground and background surface estimations. In this stage, both grayscale and binarized versions of the registered verso are used to guide removal of bleedthrough from the foreground estimation. If the pixel in the recto grayscale image is less than or equal to the corresponding pixel in the grayscale registered reverse, the corresponding pixel in the binary reverse is equal to 1, and the value of the pixel in the source grayscale image is less than a heuristically defined limit, then the value of the corresponding pixel in the foreground estimation is set to 0. Samples of this variant and a recto-verso variant from the KL thresholding family appear in Figure 1.

3 RESULTS

The data set we used for our experiments includes a total of 167 pages from three different books: two printed books of music by Luca Marenzio from 1581 and 1598, RISM M-0539 and RISM M-0582 [19], and one manuscript (handwritten book) compiled by the Alamire scriptorium in 1534, the Occo Codex. The 1581 Marenzio book contains significantly more bleedthrough than the 1598 book, and its landscape layout presented the registration process with additional challenges. The Occo Codex, where the amount of bleedthrough is comparable to the 1581 book but the layout is the same as the 1598 book, presented further challenges because most OMR systems, including our own (Aruspix [18]), are designed for printed scores rather than manuscripts.

The performance of recognition systems is traditionally presented in terms of recall and precision, but in the case of OMR, a more relevant statistic is the amount of time it will take a human to correct recognition errors. We present our results using a variant of our editing cost formula in [18], specifically

$$V = 1 - \left(\frac{{}^1\!/\!_4\, I + {}^1\!/\!_2\, S + D}{N}\right) , \qquad (5)$$

where I is the number of musical symbols wrongly inserted by the OMR process, S is the number of musical symbols incorrectly classified by the OMR process, D is the number of musical symbols missed by the OMR process, and N is the total number of symbols on the page. The quantity V represents the savings in human editing time relative to transcribing a page by hand, i.e., $V = 0.75$ would mean that only a quarter of the human labor necessary for a music digitization project without OMR would be needed after introducing the OMR system. Perfect OMR would yield $V = 1.00$.

As the results in Table 1 show, all of these algorithms are good (as one would expect considering that top performers were selected for the experiments). Excepting the special cases of symmetric and asymmetric KL thresholding with three classes, which will be discussed below, the minimum average cost saving is 71 percent (asymmetric KL thresholding with two classes on the Occo Codex). For more accurate comparisons among algorithms, we also provide the odds multipliers (model coefficients transformed with the inverse link function) from a binomial regression model

Algorithm	RISM M-0539			RISM M-0582			Occo Codex		
	Rate	Odds multiplier		Rate	Odds multiplier		Rate	Odds multiplier	
Symmetric KL: 2 classes	0.88	0.56	(0.46, 0.67)	0.87	0.70	(0.65, 0.75)	0.75	0.81	(0.74, 0.87)
Symmetric KL: 3 classes	**0.93**	**1.00**	–	**0.90**	**1.00**	–	0.17	0.05	(0.05, 0.06)
Symmetric KL: recto-verso	0.89	0.66	(0.54, 0.81)	0.88	0.78	(0.71, 0.86))	0.76	0.86	(0.79, 0.93)
Asymmetric KL: 2 classes	0.88	0.57	(0.47, 0.69)	0.87	0.69	(0.64, 0.74))	0.71	0.65	(0.60, 0.70)
Asymmetric KL: 3 classes	0.93	0.99	(0.81, 1.22)	0.90	0.99	(0.92, 1.06)	0.11	0.03	(0.03, 0.04)
Asymmetric KL: recto-verso	0.89	0.63	(0.52, 0.77)	0.88	0.75	(0.69, 0.83)	0.76	0.85	(0.78, 0.92)
Gatos et al. (31-px window)	0.87	0.52	(0.43, 0.62)	0.89	0.86	(0.80, 0.92)	**0.79**	**1.00**	–
Gatos et al. (21-px window)	0.75	0.23	(0.19, 0.27)	0.88	0.81	(0.75, 0.87)	0.77	0.89	(0.82, 0.97)
Gatos et al. (31-px window): recto-verso	0.90	0.68	(0.56, 0.83)	0.89	0.88	(0.80, 0.97)	0.76	0.83	(0.77, 0.90)
Gatos et al. (21-px window): recto-verso	0.86	0.46	(0.38, 0.56)	0.88	0.81	(0.73, 0.89)	0.75	0.80	(0.74, 0.87)

Table 1: Average savings rates over manual transcription time for each algorithm and book. The corresponding odds multipliers (inverse-link-transformed regression coefficients), given here with their 95-percent-confidence intervals, take into account the variable difficulty of OMR for the different books and are scaled relative to a baseline of the best-performing algorithm for each book. Where the confidence intervals do not overlap, the differences are statistically significant.

over all pages in our experiment [4, 14], which take into account the fact that some books in the set are more difficult than others, a fact which is highly statistically significant ($p < 0+$). The baseline for the odds multipliers is set relative to the highest-performing algorithm for each book; thus they range from 0 to 1 and represent the proportion of the maximum recognition performance produced by any given algorithm. The 95-percent confidence intervals for the odds multipliers are also included. Where these intervals overlap, there is no statistically significant difference between the two algorithms.

The first point to note about the results is that although the symmetric KL thresholding family seems to perform slightly better than the asymmetric KL thresholding family on average, there is no statistically significant difference between them. For the printed music (RISM M-0539 and RISM M-0582), the three-class variants of KL thresholding are the clear winners. In the case of the Occo Codex, however, this pair of algorithms performs abysmally. The reason is clear from Figure 1f: in the presence of uneven lighting and variable ink density (e.g., from a quill pen rather than a printing press) these algorithms suffer a strong tendency to overclean. Curiously, our experiments showed that there was no statistically significant difference between the performance of the originally published version of the symmetric *two-class* KL algorithm [2], which contains an error in the normalization for mean foreground and background values, and the *three-class* variants of KL thresholding in our experiments. This unexpected similarity was true both for the Marenzio prints, for which these algorithms perform extremely well, and for the Occo Codex, for which these algorithms overclean and thus perform extremely poorly.

Although the global thresholds chosen by the two-class KL algorithms yield fair results, our adaptive algorithm set, based on [10], performs the best on the Occo Codex. In contrast to [4], for which the ideal window size was 21 pixels, a window size of 31 pixels works best on this manuscript. Looking at the performance of these algorithms for the print books, however, the preference for a 31-px window is not always statistically significant, which is in keeping with our earlier finding that choosing a single ideal window size for adaptive thresholding algorithms is impossible [4].

Incorporating information from the verso did not help the KL thresholding algorithms significantly. Although there are small improvements in the savings rate for all books in our data set, the confidence intervals for two-class KL thresholding and recto-verso KL thresholding all overlap considerably. For the Gatos-style algorithms, however, the registration does make a difference for RISM M-0539, the book with the most severe bleedthrough in our data set, although the improvement is only statistically significant for a window size of 21 pixels. Considering the computational resources necessary to register images, however, the improvements are unlikely to be worth the computational cost for all but the most bleedthrough-ridden OMR projects.

4 SUMMARY AND FUTURE WORK

On account of their computational speed and ease of implementation, we would recommend our new symmetric three-class KL thresholding algorithm for printed music and Gatos et al.'s algorithm for documents with uneven lighting or ink distribution, such as most manuscripts. Several techniques have also been presented in the literature for treating uneven lighting (e.g., [21]), and it is worth exploring whether correcting for uneven lighting before thresholding with a three-class KL algorithm would temper the overcleaning problem.

A very different technique for document binarization has also been published recently, based on independent component analysis (ICA) [22]. Early work with ICA for binarization was restricted to color images because multiple image channels are required for ICA to be meaningful. The more recent work uses recto-verso registration to yield the two-channel minimum necessary for ICA. Because this algorithm is so markedly different from traditional approaches, we have not yet had an opportunity to compare it to our own algorithms. We expect that the presence of staves on both sides of the page will be too strong a violation of the independence assumption for ICA to work well for music images, but because the authors claim that the technique is more sensitive to the quality of registration than the content of the recto and verso pages, we are planning to explore their technique on our database of music images in the near future.

5 ACKNOWLEDGEMENTS

We would like to thank the Canada Foundation for Innovation and the Social Sciences and Humanities Research Council of Canada for their financial support. We would also like to thank Marnie Reckenberg, Tristan Matthews, and Jane Hatter for their contributions to the project.

6 REFERENCES

[1] Baird, H. S. The state of the art of document image degradation modeling. In *Proceedings of the 4th IAPR International Workshop on Document Analysis Systems* (Rio de Janeiro, Brazil, 2000), pp. 1–16.

[2] Brink, A. D., and Pendock, N. E. Minimum cross-entropy threshold selection. *Pattern Recognition 29*, 1 (1996), 179–88.

[3] Bruder, I., Finger, A., Heuer, A., and Ignatova, T. Towards a digital document archive for historical handwritten music scores. In *Digital Libraries: Technology and Management of Indigenous Knowledge for Global Access*, T. M. T. Sembok, H. B. Zaman, H. Chen, S. Urs, and S. H. Myaeng, Eds., vol. 2911 of *Lecture Notes in Computer Science*. Springer, Berlin, 2003, pp. 411–14.

[4] Burgoyne, J. A., Pugin, L., Eustace, G., and Fujinaga, I. A comparative survey of image binarisation algorithms for optical recognition on degraded musical sources. In *Proceedings of the 8th International Conference on Music Information Retrieval* (Vienna, Austria, 2007), pp. 509–12.

[5] Castro, P., Almeida, R. J., and Pinto, J. R. C. Restoration of double-sided ancient music documents with bleed-through. In *Progress in Pattern Recognition, Image Analysis, and Applications*, vol. 4756 of *Lecture Notes in Computer Science*. Springer, Berlin, 2007, pp. 940–49.

[6] Cook, N. Towards the compleat musicologist? Invited talk, 6th Annual Conference on Music Information Retrieval, London, England, September 2005.

[7] da Silva, J. M. M., Lins, R. D., and da Rocha, V. C. Document engineering (DE): Binarizing and filtering historical documents with back-to-front interference. In *Proceedings of the 2006 ACM Symposium on Applied Computing* (Dijon, France, 2006), ACM Special Interest Group on Applied Computing, ACM Press, pp. 853–58.

[8] Dano, P. Joint restoration and compression of document images with bleed-through distortion. Master's thesis, University of Ottawa, Ottawa, Canada, 2003.

[9] Fadoua, D., Le Bourgeois, F., and Emptoz, H. Restoring ink bleed-through degraded document images using a recursive unsupervised classification technique. In *Document Analysis Systems VII*, H. Bunke and A. L. Spitz, Eds., vol. 3872 of *Lecture Notes in Computer Science*. Springer, Berlin, 2006, pp. 38–49.

[10] Gatos, B., Pratikakis, I., and Perantonis, S. J. An adaptive binarization technique for low quality historical documents. In *Document Analysis Systems VI*, S. Marinai and A. Dengel, Eds., vol. 3163 of *Lecture Notes in Computer Science*. Springer, Berlin, 2004, pp. 102–13.

[11] Johansson, M. Image registration with simulated annealing and genetic algorithms. Master's thesis, Kungliga Tekniska Högskolan, Stockholm, Sweden, 2006.

[12] Leedham, G., Varma, S., Patankar, A., and Govindarayu, V. Separating text and background in degraded document images – a comparison of global threshholding techniques for multi-stage threshholding. In *Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition (IWFHR'02)* (2002).

[13] Li, C. H., and Lee, C. K. Minimum cross-entropy thresholding. *Pattern Recognition 26*, 4 (1993), 617–25.

[14] McCullagh, P., and Nelder, J. A. *Generalized Linear Models*, 2nd ed. Chapman and Hall, London, 1989.

[15] Mendel, A. Some preliminary attempts at computer-assisted style-analysis in music. *Computers and the Humanities 4*, 1 (1969), 41–52.

[16] Niblack, W. *An Introduction to Digital Image Processing*. Prentice Hall, Englewood Cliffs, NJ, 1986, pp. 115–16.

[17] Otsu, N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics 9*, 1 (1979), 62–66.

[18] Pugin, L., Burgoyne, J. A., and Fujinaga, I. Reducing costs for digitising early music with dynamic adaptation. In *Research and Advanced Technology for Digital Libraries*, L. Kovács, N. Fuhr, and C. Meghini, Eds., vol. 4675 of *Lecture Notes in Computer Science*. Springer, Berlin, 2007, pp. 471–74.

[19] Répertoire international des sources musicales (RISM). *Single Prints Before 1800*. Series A/I. Bärenreiter, Kassel, 1971–81.

[20] Sharma, G. Show-through cancellation in scans of duplex printed documents. *IEEE Transactions on Image Processing 10*, 5 (May 2001), 736–754.

[21] Shi, Z., and Govindaraju, V. Historical document image enhancement using background light intensity normalization. In *Proceedings of the 17th International Conference on Pattern Recognition* (Cambridge, United Kingdom, August 2004).

[22] Tonazzini, A., Salerno, E., and Bedini, L. Fast correction of bleed-through distortion in grayscale documents by a blind source separation technique. *International Journal of Document Analysis 10*, 1 (May 2007), 17–25.