A ROBUST BORDER DETECTION ALGORITHM WITH APPLICATION TO MEDIEVAL MUSIC MANUSCRIPTS

Yue Ouyang
Centre for Intelligent Machines
McGill University, Montréal, Canada
yue.ouyang@mail.mcgill.ca

John Ashley Burgoyne, Laurent Pugin, and Ichiro Fujinaga
CIRMMT, Schulich School of Music
McGill University, Montréal, Canada
{ashley,laurent,ich}@music.mcgill.ca

ABSTRACT

Medieval music manuscripts pose special challenges for digital processing. Their unique page layouts and sometimes extreme degradation can make it difficult even to identify the portions of an image that correspond to the musical page. This paper addresses the page identification problem for medieval documents, with natural extensions to any type of image processing that entails separating a border from the central image content. Aimed at keeping the removal process simple and fast, we present a contour-searching approach that is novel for using only the paper itself, without cues from staff lines or other musical elements.

1. INTRODUCTION

Medieval music manuscripts are among the most beautiful documents with which both musicians and music scholars can work; however, many of the features that make these documents fascinating can pose problems for digital processing. Page layouts are nonstandard; contrast can be low. Due to aging, ink may be faded or have bled through the paper, and the paper may be crumbling, stained, torn, or even burnt [6]. Moreover, when these documents are scanned or photographed, reference features such as rulers and colour bars are often present surrounding the image, background colours can be inconsistent, and illumination can vary across the page. This combination of physical and digital problems makes it difficult for computers even to locate which portions of an image correspond to the musical page, let alone the locations of music, lyrics, or other elements.

This paper focuses on border detection: separating the paper from background elements that are part of the scanning or photographic process. There are a variety of applications that require this type of elementary layout analysis, but the most common reason to do this with medieval music manuscripts would be as a precursor to information extraction such as optical character recognition (OCR) or optical music recognition (OMR). The promise of OMR for these documents is the creation of large-scale databases of music, which could then be searched for musical and textual content in addition to metadata, such as composer, title, etc. [2].

Research in document image analysis has explored several avenues. Some of the more recent advances include complex models from machine learning such as Markov and conditional random fields [4]. Other approaches are based on techniques from signal processing, including independent component analysis [8]. The classic approaches rely on run length and edge detection and are still being used and improved today, in part because of their computational efficiency [1, 3]. Because image analysis is usually employed as a precursor to OMR, music-specific approaches have tended to focus on finding staff lines [7].

In this paper, we solve the border detection problem by simplifying the layout of the image and utilizing classical edge detection. Imagine a situation in which the pen strokes on a score sheet have totally faded away. The local contrast within the now blank paper is always lower than the local contrast near the paper boundary. Edges with high local contrast are extracted, including the contour of the blank score paper. Colour bars and rulers are normally placed such that they do not overlap the paper, and thus, once a paper-enclosing edge has been detected, these reference elements are no longer obstacles. However, because of the degradation of these very old documents, local contrast loss usually interrupts this contour, and thus, a reconnecting process is necessary. Our algorithm is robust with respect to curved contours of score sheets because it does not depend on line tracking. Moreover, because no explicit OMR stage is required, the performance is invariant to the appearance of musical symbols and reference elements around the sheet.

2. BORDER REMOVAL

The main process of the border removal algorithm is illustrated in Figure 1. The algorithm includes steps for preprocessing, background estimation, boundary detection, and boundary reconstruction.

2.1. Preprocessing

For the purpose of OMR, music documents have usually been scanned at a high resolution and converted to grayscale. For border removal, however, the content of the documents

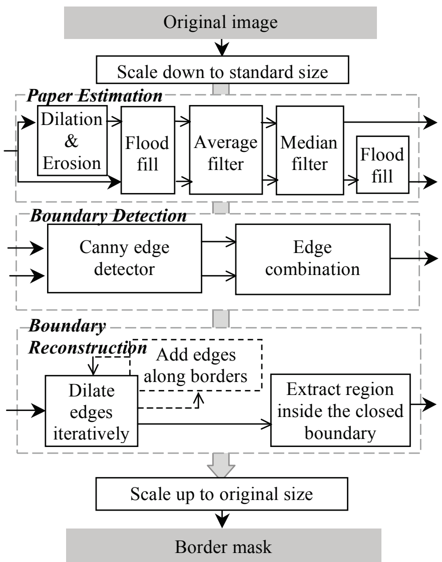

Figure 1. Block diagram of our border-removal algorithm. The upper line reflects the version of the image with extra smoothing; the lower line reflects the version of the image with more detail. The final border mask can be applied directly to the image or retained for other types of processing.

can be ignored, as no recognition process is involved. Thus, we scale the images to a standard low resolution. After the border of the paper has been located, the border mask is scaled back to its original size.

2.2. Background Paper Estimation

2.2.1. Levels of smoothness

Our algorithm utilizes an approximation of blank paper rather than the original musical score. This approximation works like a filter in scale space. Small details like the pen strokes on the score page are filtered out, but the large-scale structure, in particular, the shape of the paper, is maintained. With historical documents, however, the shape of the paper can also contain finer structures: jagged contours, texture arising from non-uniform illumination inside the page, and stains at the edges. It is difficult for a single filter to separate the finer structures included in the paper contour from those belonging to the pen strokes.

In our method, two estimations of the paper contour are obtained at different levels of smoothness, one with a high level of smoothness discards most details and contains only large-scale structures in the image, while another with a lower level of smoothness preserves finer structures. This approach is similar to the "pyramid" approaches in [5]. The smoother version gives an approximate but high-confidence estimation of the shape of the paper, while the sharper version contains details about the paper contour that can correct blur in the original estimation of the paper shape. Ultimately, edges are detected using each of these two estimation and are combined into a single contour.

The two levels of smoothness are obtained from slightly different routines. In addition to the main stage of paper estimation, extra processing may be used before or after the main stage depending on the desired level of smoothness. For the smoother estimation, an extra grayscale morphological operation, the so-called "close" operation, is employed at the beginning of the process, which largely suppresses the local variance in contrast along the contours of pen strokes. In order to compute the less smooth estimation, a flood-fill process is applied at the end of estimation so that more details of local contrast are preserved. Figure 2 shows (a), a sample image, and (b), a portion of it also shown after each of the two levels of smoothing, (c) and (d).

2.2.2. Pen stroke removal

In this section, we explain the algorithm that removes the pen strokes from the music sheet. Such strokes are recoloured with the local grayscale intensity of the surrounding paper, and hence, theoretically blank paper is obtained. Pen strokes always have lower intensity then the paper (i.e., they are darker), and they are almost always are isolated from the image border. With these properties, the strokes satisfy the definition of "holes" in grayscale morphology. Flood-fill [9], a common morphological reconstruction approach to hole-filling, can therefore be applied. Figure 3 shows an example from one of our test pages. Flood-fill removes isolated pen strokes and stains effectively. However, the potential contour of the music paper is preserved so long as either the background of a dark-background image extends all the way to the image border or the contour of the paper against a bright background is strong enough to obtain high variance of intensity.

2.2.3. Paper smoothness

Continuing with the paper estimation process, the image is further smoothed by an average filter and a median filter in order to eliminate edges irrelevant to the contour of the score paper. Because the pen strokes are removed by flood-fill rather than by interpolation, each pen stroke is uniformly filled with the lowest intensity of the paper around its contour. The local contrast around the trace of a pen stroke trace varies. To suppress such variance and obtain a smoother estimation inside the paper, average filtering is applied. Because of the relatively high local contrast, the contour of the paper is preserved during this blurring process.

Unlike the condition discussed above, flood-fill and average filtering fail to bring the intensity of traces of pen strokes up to the intensity of the paper when the stroke is not entirely isolated, i.e., the lowest intensity of the paper around the pen stroke is much lower than the average. Such

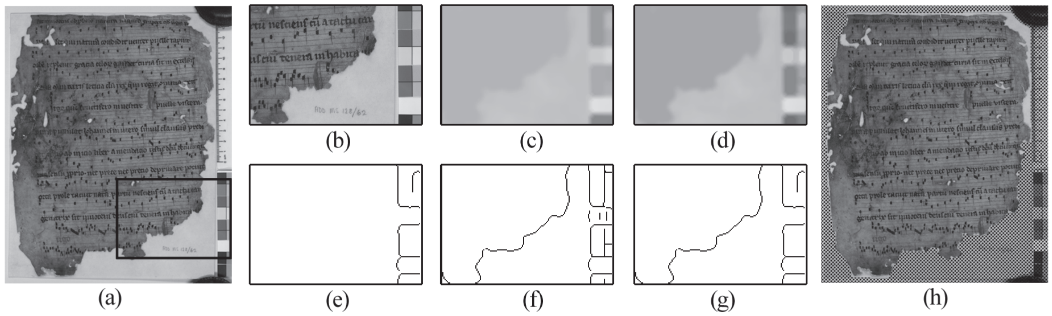

Figure 2. a) Original image. b) Portion of the original image. c) Result of background paper estimation with further smoothness. d) Result of background paper estimation without further smoothness. e) Edge map extracted from (c). f) Edge map extracted from (d). g) Final edge map after combination. h) Border mask applied to the original image.

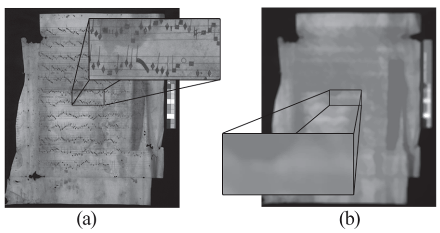

Figure 3. Background estimation. a) Original image. b) Background image obtained using flood fill and blurring.

defects can be remedied by median filtering, which discards extreme values and replaces them with the median value inside the filter window.

2.3. Boundary Detection and Reconstruction

In this section, we describe our approach to extract the boundary of the score paper. First, the Canny edge detector is introduced, and then we explain the edge combination strategy that combines edges extracted from different levels of smoothness. From these extracted edges, the boundary is reconstructed without supervision.

Due to degradation, local contrast varies along the boundary of elements in the image, and it is difficult to extract a closed contour surrounding the paper. Edges with high local contrast (strong edges), which are located by detectors with high thresholds, are often stymied by local contrast loss, but edge detectors with low thresholds are easily fooled by non-uniform illumination. This problem can be remedied by including weak edges that are near to strong ones; the Canny edge detector [1] is a good candidate for this situation. Rather than using a single threshold, the Canny edge detector employs a pair of thresholds, high and low, for strong and weak edges respectively. In our experiment, a higher set of thresholds is applied to the smoother paper estimation containing major structures, and a lower set is applied to the other estimation of a low level of smoothness. The two outputs of edge detection are shown in Figure 2 (e) and (f).

After Canny edge detection, two edge maps are available and are combined into one. The first map, derived from the smoother estimation, is clean, and most edges in it are the contour fragments of major structures in the image. The second map contains potential fragments that may make up the gaps along contours as well as contours that might be lost in the first map. The final edge map is computed by adding salient edges from the second map to the first map. Assuming that long and relatively straight edges are more likely to be part of the contour, edges containing such segments are considered to be salient edges. In our experiment, the defining feature of a salient edge is the ratio of the length of the longest side of its bounding box to that of the image. A threshold strategy is then applied to select salient edges in the second map, but unfortunately, there is no strict rule to determine this threshold; we have tuned it manually. Figure 2 (g) shows the final edge map after combination.

Although fragments of the paper contour are gathered by edge detection as described above, they do not always form a closed boundary and thus must be reconnected. Moreover, the physical score sheet may run past the scanned region, resulting in undetected edges at the border. We use isotropic morphological techniques to reconstruct the boundary from the fragments. A dilation process is iterated until an enclosed contour surrounding the image centre emerges. The skeleton of this contour is extracted as the paper boundary.

3. EXPERIMENTS AND RESULTS

Fifty images with different types of degradation were chosen from the Digital Image Archive of Medieval Music (DIAMM) as our testing set.[1] The set comprised historical manuscripts of particularly poor quality so as to encompass a broad spectrum of challenges confronting OMR for early documents. These documents had non-uniform illumination

[1] http://www.diamm.ac.uk/

at various levels, and most of them have irregular boundaries that are either locally fuzzy or non-rectangular (or both). Moreover, the backgrounds on which the music scores were placed had different colours, and sometimes there were even multiple background colours within a single image. Reference elements (colour bars and rulers) also varied considerably in both their position and overall appearance.

Unlike other processes in OMR, it is extremely difficult to evaluate border removal algorithms quantitatively. It is impractical to define the boundary of every page in a testing set at the pixel level when the boundaries are so irregular. On the other hand, one can easily tell by human inspection if border removal works successfully. Hence, we categorize the results coarsely into four sets, "excellent," "good," "acceptable," and "wrong." A result is labelled as "excellent" when the music score is well extracted and the boundary of paper is preserved completely. When the estimated boundary slightly varies from the ground truth but does not harm the pen stroke or include reference elements, it is categorized as "good," e.g., Figure 2 (g). A removal is "acceptable" if small errors occur, such as including small bits of reference elements (colour bars and rulers) or trimming out several pen strokes. When the errors on removal are serious, the result is labelled as "wrong." In our experiment, a set of fixed global parameters was applied on the whole testing set. In the end, 40 percent of the results were labelled as "excellent," 26 percent as "good," 24 percent as "acceptable," and 10 percent as "wrong."

Several points are worthy to note about the results. First, despite the selection of parameters, it was impossible to achieve "excellent" for images with highly fuzzy score boundaries because the fuzzy details are lost during dilation in boundary reconstruction. Second, by looking into the 10 percent rate of failure, the reasons can be abstracted into two general categories: local contrast loss and isotropic dilation. The paper-shape estimation algorithm performs badly on images with low-contrast contours. Due to this local contrast loss, the gaps to reconnect the paper contour can be much wider than the distance to spurious edges, and thus during isotropic dilation, an incorrect boundary is found.

4. SUMMARY AND FUTURE WORK

In this paper, we propose an approach to border removal for historical documents in which pen strokes on the paper are treated as "noise." With edge-preserving denoising, blank paper is estimated, and we show how it can be extracted from the background. Then, high local contrast detection locates the candidate contour fragments. In the reconnection stage, the contour fragments grow into a free-form closed boundary, enabling our algorithm to handle music scores of irregular shape.

When OMR has improved to a level such that it can recognize the symbols in medieval manuscripts, we would also like to replace the subjective evaluation we used for our experiment with a quantitative, goal-directed evaluation procedure based on OMR (see [7]).

The work is beneficial even for researchers outside OMR. We are using the system as part of an OCR system for medieval documents, again as part of a larger project for digital libraries. As mentioned above, the algorithm is also suitable for other music-related images such as record labels or album art, and its computational efficiency suggests potential uses on mobile devices and for processing multimedia.

5. ACKNOWLEDGEMENTS

We thank SSHRC and the CFI for their generous support.

6. REFERENCES

[1] J. Canny, "A computations approach to edge detection," *IEEE Trans. PAMI*, vol. 8, pp. 679–714, 1986.

[2] G. S. Choudhury, T. DiLauro, M. Droettboom, I. Fujinaga, B. Harrington, and K. MacMillan, "Optical music recognition system within a large-scale digitization project," in *Proc. Int. Conf. Mus. Inf. Ret.*, 2000.

[3] J. Freixenet, X. Muñoz, D. Raba, J. Martí, and X. Cufí, "Yet another survey on image segmentation: Region and boundary information integration," in *Comp. Vis. — ECCV 2002*. Berlin: Springer, 2002, pp. 408–22.

[4] X. He, R. Zemel, and M. A. Carreira-Perpiñán, "Multiscale conditional random fields for image labeling," in *Proc. IEEE Conf. Comp. Vis. and Pat. Rec.*, 2004.

[5] R. Manmatha and J. L. Rothfeder, "A scale space approach for automatically segmenting words from historical handwritten documents," *IEEE Trans. PAMI*, vol. 27, pp. 1212–25, Aug 2005.

[6] J. C. Pinto, P. Vieira, M. Ramalho, M. Mengucci, P. Pina, and F. Muge, "Ancient music recovery for digital libraries," in *Proc. Eur. Conf. Res. and Adv. Tech. for Dig. Lib.*, 2000, pp. 24–34.

[7] L. Pugin, J. A. Burgoyne, and I. Fujinaga, "Goal-directed evaluation for the improvement of optical music recognition on early music prints," in *Proc. ACM/IEEE Joint Conf. Dig. Lib.*, Vancouver, Canada, 2007, pp. 303–04.

[8] A. Tonazzini, E. Salerno, and L. Bedini, "Fast correction of bleed-through distortion in grayscale documents by a blind source separation technique," *Int. J. Doc. Anal.*, vol. 10, no. 1, pp. 17–25, 2007.

[9] L. Vincent, "Morphological grayscale reconstruction in image analysis: Application and efficient algorithms," *IEEE Trans. Im. Proc.*, vol. 2, no. 2, pp. 176–201, 1993.