IEICE Electronics Express, Vol.9, No.7, 609–615

A novel staff removal method for printed music image

Farshad Tajeripour[a] **and Mahmood Sotoodeh**[b]
School of Electrical and Computer Eng., Shiraz University, Shiraz, Iran
a) tajeri@shirazu.ac.ir
b) msetoodeh@cse.shirazu.ac.ir

Abstract: The first phase in optical music recognition (OMR) is staff-removal. The goal of this phase is to remove staffs and keep musical symbols with high precision, which is very important for next phases in OMR including segmentation and classification. In this paper a novel staff-removal method is proposed. In this method we used run-length coding to compress the input image and then the staff-lines are detected and removed using two common features, namely the thickness of lines and the space between them. Afterwards, The nonstaff objects such as long objects, dynamic wage and destroyed texts are removed using projection and morphological operations.

Keywords: staff-removal, staff-line, run-length, OMR, morphology, projection
Classification: Other communication hardwares

References

[1] P. Bellini, I. Bruno, and P. Nesi, "Optical music sheet segmentation," *Proc. First International Conference on Web Delivering of Music*, pp. 183–190, Nov. 2001.
[2] L. Pugin, "Optical music recognition of early typographic prints using hidden Markov models," *Proc. 7th International Conference on Music Information Retrieval (ISMIR)*, Citeseer, pp. 53–56, 2006.
[3] R. Randriamahefa, J. Cocquerez, C. Fluhr, F. Pepin, and S. Philipp, "Printed music recognition," *Proc. Second International Conference on Document Analysis and Recognition*, pp. 898–901, Oct. 1993.
[4] I. Fujinaga, "Staff detection and removal," *Visual Perception of Music Notation: On-Line and Off Line Recognition*, ed. S. E. George, pp. 1–39, 2004.
[5] F. Toyama, K. Shoji, and J. Miyamichi, "Symbol recognition of printed piano scores with touching symbols," *International Conference on Pattern Recognition*, vol. 2, pp. 480–483, 2006.
[6] H. Miyao and Y. Nakano, "Note symbol extraction for printed piano scores using neural networks," *IEICE Trans. Inf. & Syst.*, vol. E79-D, no. 5, pp. 548–554, 1996.
[7] F. Rossant and I. Bloch, "Robust and adaptive omr system including fuzzy modeling, fusion of musical rules, and possible error detection," *EURASIP J. Appl. Signal Process.*, vol. 2007, p. 160, Jan. 2007.

DOI: 10.1587/elex.9.609
Received January 10, 2012
Accepted February 14, 2012
Published April 01, 2012

IEICE Electronics Express, Vol.9, No.7, 609–615

[8] J. Roach and J. Tatem, "Using domain knowledge in low-level visual processing to interpret handwritten music: An experiment," *Pattern Recognition*, vol. 21, no. 1, pp. 33–44, 1988.
[9] J. Mahoney, "Automatic analysis of music score images," Ph.D. dissertation, Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, 1982.
[10] K. Reed and J. Parker, "Automatic computer recognition of printed music," *Proc. 13th International Conference on Pattern Recognition*, vol. 3, pp. 803–807, Aug. 1996.
[11] H. Miyao and M. Okamoto, "Stave extraction for printed music scores using dp matching," *J. Advanced Computational Intelligence and Intelligent Informatics*, vol. 8, no. 2, pp. 208–215, 2004.
[12] A. Dutta, U. Pal, A. Fornes, and J. Llados, "An efficient staff removal approach from printed musical documents," *International Conference on Pattern Recognition*, pp. 1965–1968, 2010.
[13] C. Dalitz, M. Droettboom, B. Pranzas, and I. Fujinaga, "A comparative study of staff removal algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, pp. 753–766, 2008.
[14] J. dos Santos Cardoso, A. Capela, A. Rebelo, C. Guedes, and J. P. da Costa, "Staff detection with stable paths," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, pp. 1134–1139, 2009.
[15] A. Fornes, A. Dutta, A. Gordo, and J. Llados, "CVC-MUSCIMA: A Ground-Truth of Handwritten Music Score Images for Writer Identification and Staff Removal," *International Journal on Document Analysis and Recognition*, in press (DOI: 10.1007/s10032-011-0168-2), 2011.

1 Introduction

All musical documents contain a set of parallel horizontal lines which are called staff or stave. According to the type of a musical document the number of lines per staff is different. For example in tablature documents each staff has six horizontal lines, but in modern musical documents each staff has five horizontal lines. Two important features of staffs are Staff-line height and Staff-space height (see Fig. 1). Staff-line height is the thickness of the lines in the staff and Staff-space height is the space between the two lines of the staff. The first step of optical music recognition is staff-removal. Although it is a very important step, but there are suggested algorithms that do not need to remove staffs [1, 2]. The two musicals documents that are the focus of most of the researchers are printed and handwritten musical documents. Since the staff-lines in handwritten musical documents are not always

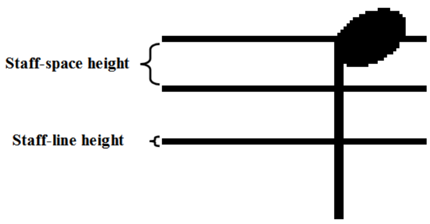

Fig. 1. Staff-line height and staff-space height.

DOI: 10.1587/elex.9.609
Received January 10, 2012
Accepted February 14, 2012
Published April 01, 2012

straight (horizontal or parallel), staff detection and removal could become a complicated process.

There are different algorithms for staff-removal. Some of them use local maximum in horizontal projection of the black pixels of the image [3, 4] or in vertical projection [5] with Hough Transform [6]. Assuming these lines are straight and horizontal, then local maximum detect the lines. This method is based on a Line Adjacency Graph (LAG). LAG looks for potential sections of line: sections that satisfy criteria related to aspect ratio, connectedness and curvature. For gaining better performance, there are some methods that use combination of projection techniques [1, 7].

Fujinaga used a set of image processing techniques, including run-length coding (RLC), connected component analysis, and projection [4]. After performing a run-length coding on the image to find the thickness of staff-lines and space between the staff-lines, any vertical black run that is greater than twice the staff-line height (i.e., staff-line thickness) is removed from the original image. Then connected components are scanned and any component whose width is less than staff-space height is removed.

In the literature, there are other proposed algorithms for finding and removing staff-lines. Some of them used line tracing to find staff-lines [8]. Mahoney [9] used rule-based classification of thin horizontal line segments. In [10] Reed and Parker proposed a method for grouping of vertical columns based on their spacing, thickness and vertical position on the image. Miyao and Okamoto used staff segmentation [11]. In this method, Segments are horizontally linked based on DP matching. Dalitz et al. in [13] improved the methods of Miyao and Okamoto. Dutta et al. proposed a method in which a staff-line segment is considered as a horizontal linkage of vertical black runs with uniform height, and neighboring properties are used to validate it [12]. Cardoso et al. proposed a method based on graph theoretic approaches [14]. This algorithm uses the image as a graph, where the staff-lines result as connected paths between the two lateral margins of the image. A staff-line can be considered as a connected path from the left side to the right side of the music score. As staff-lines are almost the only extensive black objects on the music score, the path to look for is the shortest path between the two margins if paths (almost) entirely through black pixels are favoured.

This paper is organized as follows. Section 2 describes the proposed method. Section 3 reports the results of our algorithm in details. Finally in Section 4 the concluding statements of the paper and future works are given.

2 Proposed methodology

Our method is based on run-length coding. Assume that we have an original image as is depicted in Fig. 2 a. Let runs be a sequence of pixels with the same color (or other characteristics). At first, the vertical run-length coding method is used to have a compact representation of the image. Then, two common features including staff-line height (SLH) and staff-space height (SSH) is employed to remove the staff-lines. SLH is the most frequent black-

DOI: 10.1587/elex.9.609
Received January 10, 2012
Accepted February 14, 2012
Published April 01, 2012

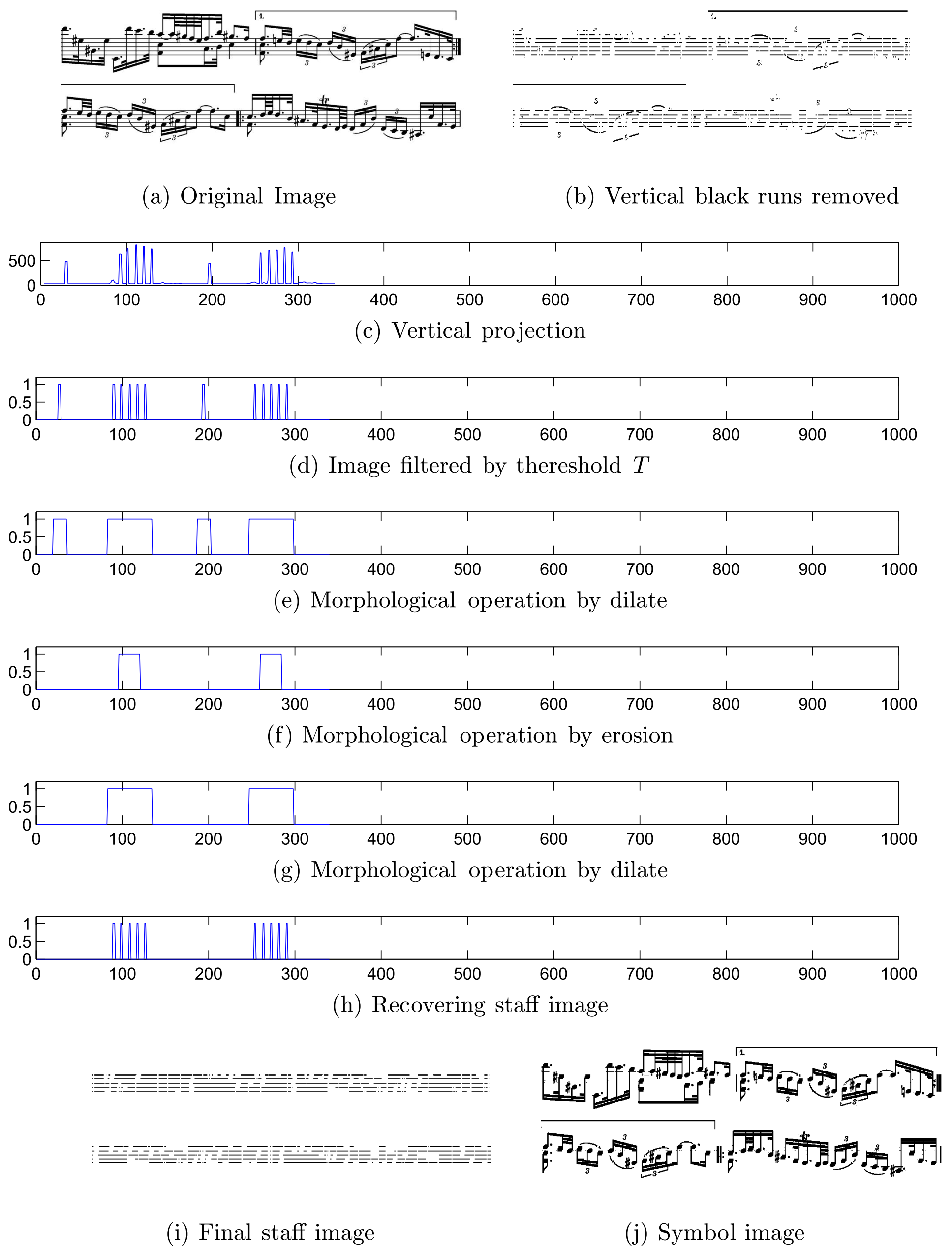

Fig. 2. Steps of the proposed algorithm applied on a sample image.

runs and SSH is the most frequent white-runs. After estimation the SLH and SSH, we use a coarse-to-fine approach for removing the staff-lines from the image. In the coarse step, any vertical black runs whose heights are more than twice the SLH are removed from the image (Fig. 2 b). In spite of this filtering, some non-staff objects such as long objects, dynamic wages, destroyed texts and noises remain in the resulting image. This is because the thickness of these objects is similar to the thickness of the staff-lines. To eliminate these objects we propose a fine non-staff removal step based on vertical projection. The proposed approach consist of 6 steps. At first, the vertical projection is performed on the previous result image. It produces an array with the length of the image height. Each element of the array shows

DOI: 10.1587/elex.9.609
Received January 10, 2012
Accepted February 14, 2012
Published April 01, 2012

the number of black pixels in its corresponding row in the image. Let V be the vertical projection on the complement of the binary image calculated by:

$$v_0[r] = \sum_{c=1}^{W} NOT(S_c(r,c)) \qquad \forall r = 1 \ldots H \tag{1}$$

where S_c is the resulting image of the coarse non-staffremoval step in which S_C(r c) denotes the pixel in row r and column c. Also W and H illustrate the width and the height of the image S_c, respectively. The vertical projection of the image given in Fig. 2 b is shown in Fig. 2 c. It is evident that the staff-lines create a long pick in the vertical projection (v). Primitively, we remove the peaks whose height is less than a threshold as:

$$v_1[r] = \begin{cases} 1 & \text{if } v_0[r] < T \\ 0 & otherwise \end{cases} \qquad \forall r = 1 \ldots H \tag{2}$$

where T is the pre-defined threshold which can be set as a ratio of W. We set this threshold to $W/8$. This operation eliminates the non-staff small object such as noise and symbols which were not removed in the coarse step. The output of this operation is presented in Fig. 2 d. However, there are some other musical symbols such as slurs, ties, dynamic wages and any long objects which produce a long pick in the vertical projection. This fact is obvious in the Fig. 2 d. In order to remove these objects, we propose a morphological-based schema as follows. First, a dilation operation is performed on the vertical projection. This operation connects the peak of the staff-lines together and makes a wide peak for each five staff-lines, as shown in Fig. 2 e. The corresponding peaks of the non-staff objects are distinguishable in this figure. Afterwards, we perform an erosion operation on the dilated vertical projection. As the result, the peaks of the non-staff objects are removed while the peaks of the staff-lines become thinner (Fig. 2 f). After that, a dilate operation is again performed to retrieve the corresponding peaks of each five staff-lines without extra objects (Fig. 2 g). The mentioned morphological operations can be summarized as:

$$v_2 = dilate(erode[dilate(v_1, M), M], M) \tag{3}$$

where, the parameter M (i.e. structuring element object) is an all-ones vector of length $3 \times SSH$. Afterward, the result of the previous step (v_2) is conjugated with the binary-vertical projection (v_1) as:

$$v_3 = AND(v_1, v_2) \tag{4}$$

The vector v_3 includes some peaks correspond to the staff-lines (Fig. 2 h). Now, we can utilize the vector v_3 to remove all non-staff objects in S_c:

$$S_F = NOT(AND(NOT(S_c), V_3)) \tag{5}$$

where V_3 is a matrix, which size is equal to the size of the original image. Each column of V_3 is equal to vector v_3. Fig. 2 i shows the image which just contains the staff-lines. Finally, the symbols image J can be retrieved by following equation and is demonstrated in Fig. 2 j:

$$J = NOT(XOR(S_F, I)) \tag{6}$$

DOI: 10.1587/elex.9.609
Received January 10, 2012
Accepted February 14, 2012
Published April 01, 2012

IEICE Electronics Express, Vol.9, No.7, 609–615

3 Experimental results

In our experiments, we have used the benchmark test sets provided by Dalitz [13]. This test set includes 32 ideal music images and a set of deformed images generated from the ideal images [12, 13]. To evaluate the performance of the proposed method, we have used the pixel based metric [13]. The results of our experiments are shown in Table I. This table reports the average Precision, Recall and the Error rate [15]. Our proposed method provide noteworthy results for all forms of deformations. We have obtained the highest recall for the ideal images, while the precision on the ideal images is higher than the average precision off all the images. Our results are also compared with some well-known methods (proposed in [4], [12] and [13]) in Table I. The precision of the proposed method is higher than our base method [4]. Our method achieved the highest recall and the lowest error rate.

The main advantage of the proposed method is its accuracy in removing long objects, printed text, slurs and ties. Most likely there exist long objects in music notation documents that have similar thickness to staff-lines. Our method effectively detects and eliminates these lines along with dynamic wages. Also, the music notation images usually include the text of poems and information about the composer. Our method performs much better than the others in detecting and eliminating these texts.

Table I. Results of the proposed method and well-known methods.

Deformation Type	Precision	Recall	Error rate
Ideal	92.89	99.83	0.024
Kunongo	90.80	98.99	0.036
Line Thickness Variation-v1	95.62	99.28	0.021
Line Thickness Variation-v2	95.42	99.24	0.022
Line Y Variation-v1	91.46	97.77	0.048
Line Y Variation-v2	91.41	97.87	0.048
White Spickles	93.83	98.54	0.026
Typeset Emulation	90.51	99.62	0.047
Average of the proposed method	92.73	**98.89**	**0.034**
Fujinaga [4]	88.95	92.40	0.099
Dalitz et al. [13]	**98.77**	87.59	0.064
Dutta [12]	95.98	96.22	0.037

4 Conclusion and future work

In this paper we proposed a method to remove the staff-lines from the musical images. First of all, objects having heights more than a certain threshold are removed. The threshold is based on staff-line height. Then, a projection technique is used to detect the staff-lines. Afterwards, the remained objects that are not staff-lines are removed using morphological operations. At the end, the detected staff-line image is mapped to the original image, so all objects but the staff-lines could be detected. The proposed method, in spite

DOI: 10.1587/elex.9.609
Received January 10, 2012
Accepted February 14, 2012
Published April 01, 2012

ELectronics EXpress

of its simplicity, is shown to perform better than the well-known methods. Moreover, the proposed method is less time consuming.

In future works we are going to apply our methods on deformed and degraded images. Major applications of our technique include producing music Braille and also writer identification. Moreover, the algorithm can be used for restoring damaged ancient images.

DOI: 10.1587/elex.9.609
Received January 10, 2012
Accepted February 14, 2012
Published April 01, 2012