An Effective Staff Detection and Removal Technique for Musical Documents

Bolan Su[1 2 *], Shijian Lu[2 +], Umapada Pal[3 †] and Chew Lim Tan[1 *]
[1]*Department of Computer Science,School of Computing,National University of Singapore*
Computing 1, 13 Computing Drive, Singapore 117417
[2]*Department of Computer Vision and Image Understanding,Institute for Infocomm Research*
1 Fusionopolis Way, #21-01 Connexis, Singapore 138632
[3]*CVPR Unit, Indian Statistical Institute*
203 B. T. Road, Kolkata, India 700108
*{*subolan,tancl*}*@comp.nus.edu.sg,[+]*slu@i2r.a-star.edu.sg,*[†]*umapada@isical.ac.in*

***Abstract*—Musical staff line detection and removal techniques detect the staff positions in musical documents and segment musical score from musical documents by removing those staff lines. It is an important preprocessing step for ensuing the Optical Music Recognition tasks. This paper proposes an effective staff line detection and removal method that makes use of the global information of the musical document and models the staff line shape. It first estimates the staff height and space, and then models the shape of the staff line by examining the orientation of the staff pixels. At last the estimated model is used to find out the location of staff lines and hence to remove those detected staff lines. The proposed technique is simple, robust, and involves few parameters. It has been tested on the dataset of the recent staff removal competition [1] held under the International Conference of Document Analysis and Recognition(ICDAR) 2011. Experimental results show the effectiveness and robustness of our proposed technique on musical documents with various types of deformations.**

***Keywords*-Optical Music Recognition; Musical Staff; Staff Line Segmentation; Staff Line Removal; Staff Line Shape Modeling**

I. Introduction

Staff detection and removal technique aims to detect and remove the staff pixels and retain the musical notes in the musical documents. In the western musical documents, the musical staffs are groups of parallel lines that are used to determine the pitch. However, these staff lines are overlapped with the musical score symbols and affect the performance of Optical Musical Recognition(OMR) systems [2]. Effective staff detection and removal techniques are very important for musical document analysis.

The most significant characteristics of staffs are the uniform thickness of staff lines and uniform space between staff lines, that are illustrated in Figure 1. All of the staff removal algorithms make use of these features. However, staff detection and removal is a difficult problem due to the following two aspects. First, the real musical documents may have different kinds of degradation such as warping and wrinkle, that deforms the staff lines. So a robust staff removal method should deal with staff lines that are not purely straight, continuous, equal-thickness, equal-distance, as illustrated in Figure 2. Second, the staffs are intersected with musical notes. It is not easy to extract the staff pixels without affecting the musical notes. The musical staff removal competition [1] held under the ICDAR 2011 shows recent effort on this issue by testing the state-of-the-art musical staff removal techniques with different kinds of deformed musical documents.

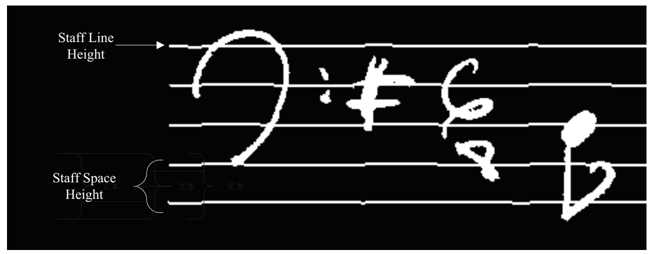


Figure 1. Staff line thickness and staff space height

Many staff line detection and removal techniques have been proposed and staff line detection is usually considered as the first step of staff removal. Horizontal and vertical projection [3], [4] is a straightforward way to detect the staff lines. Other approaches including line tracking [5], [6], staff segments [7], [8], Skeleton [2] and stable path [9] are also used for detecting staff positions. After the staff line positions are located, the music score pixels are extracted using some criteria. The staff removal algorithms can be categorized into several groups, including Line Tracking, Vector Field, Runlength and Skeletonization [2]. This paper presents a robust staff detection and removal technique. One of distinct characteristics of our proposed technique is that it makes use of the global information of the musical document and models the staff line shape. Experiments are conducted over the recent musical staff removal competition dataset, and the experimental results have proven the robustness and effectiveness of our proposed technique on different kinds of deformed musical documents.

II. Proposed Method

Figure 3 shows the overall procedure of our proposed staff detection and removal technique. The proposed technique

978-0-7695-4661-2/12 $26.00 © 2012 IEEE
DOI 10.1109/DAS.2012.16

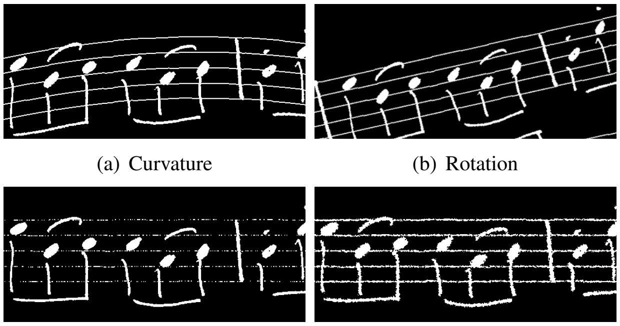


Figure 2. Several kinds of deformations of musical documents.

first estimates the staff line thickness($staff\_height$) and staff space height($staff\_space$), and uses these two features to generate an initial staff line image, in which the musical score symbols are roughly removed. Then the shape of staff line is modeled based on the information extracted on the initial staff line. The locations of staff lines are detected using the estimated staff line shape. At last, the final result is obtained by removing the detected staff lines.

The rest of this section is divided into four subsections that focus on generating the initial staff line image, modeling the staff line shape, staff line detection and staff line removal, respectively.

A. Generation of Initial Staff Line Image

The input of our staff removal method is a binary musical document with musical symbols and staffs in white. The musical staffs can be assumed more or less horizontal, parallel and uniform thickness. Then the $staff\_height$ and $staff\_space$ can be assigned as the highest frequency of vertical white-run length and black-run length, respectively.

Since the staff lines cannot have much larger thickness than the estimate $staff\_height$, we can roughly remove most of the musical score pixels by scanning the musical document column by column and removing those connected components (which can be called staff segments) that are longer than a threshold T at each column. The initial staff of Figure 3(a) is shown in Figure 3(b). As illustrated in the image, the initial staff image retains most of the musical staff pixels and erases most of the musical score pixels.

The staff line thickness must be smaller than the sum of the estimated $staff\_height$ and $staff\_space$, otherwise, the adjacent staff lines will mix together. The threshold T should be smaller than $staff\_height + staff\_space$. Considering the variation of the staff line thickness, the threshold T is set as $MIN\{2 \times staff\_height, staff\_height + staff\_space\}$ empirically.

B. Staff Line Shape Modeling

After the initial staff line image is obtained, the staff line shape can be modeled by examining the orientation of the staff line pixels. Since the staves are and parallel, the staff line shape can be modeled as an array $\{O_1, O_2, \ldots, O_n\}$ with size of the image columns n. Each element of this array O_i denotes the staff line orientation at i_{th} column, where orientation denotes the change in vertical direction between the current column and the next column. So the staff line shape can be constructed exactly from the orientation array $\{O_1, O_2, \ldots, O_n\}$

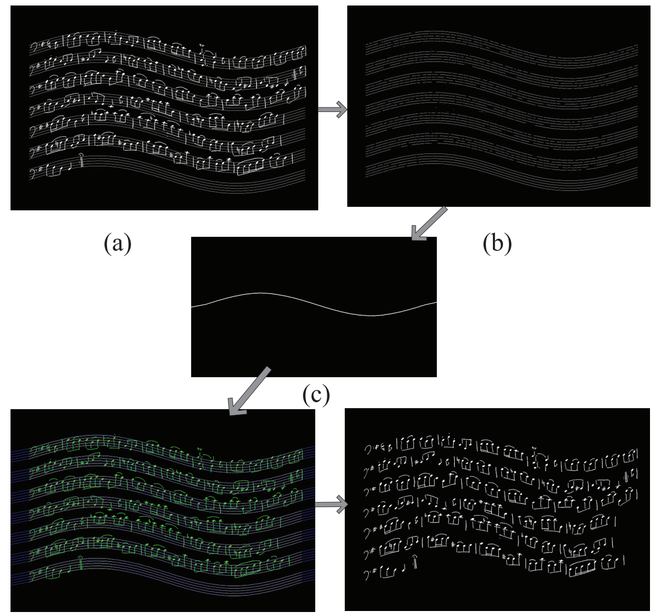


Figure 3. The procedure of our proposed method. (a) the input musical document, (b) initial staff line image, (c) the estimated staff line shape, (d) the detected staff line locations that are highlighted in white, and the musical score symbols are in green color on electronic version, (e) the final result that retained only the musical score symbol after staff line removal.

To determine the orientation array $\{O_1, O_2, \ldots, O_n\}$, we examine the initial staff image column by column. At every column, each connected component is considered as a potential staff segment. The orientation of the staff segment is defined as the row distance to the nearest staff segment in the next column, and the distance can be denoted simply as the row indexes difference between the middle pixels of the two staff segments, which is shown in Figure 4. To make the estimation of orientation more accurate, we calculate the distance between the examining segment and the nearest staff segment in the next k columns, and average them as follows:

$$Orien = \sum_{j=i+1}^{i+k} \frac{1}{j} \times (Dis(i,j)) \tag{1}$$

where $Orien$ is the needed orientation, i denotes the column index of the testing staff segments, k is a pre-defined

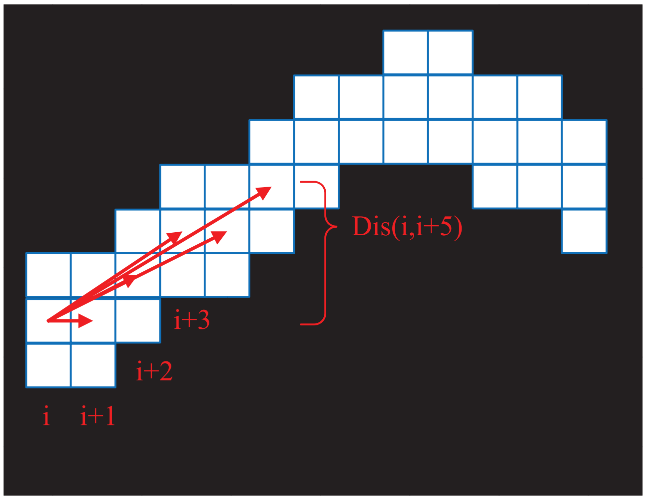


Figure 4. A fragment of a staff line, one of the white grids denotes one staff line pixel. The orientation of a staff segments is determined by the distance between it and the corresponding staff segments in the next columns. The distance is defined as the difference of the row indexes of the two staff segments.

parameter that controls how many columns are used to calculate the orientation. and $Dis(i, j)$ denotes the distance of the middle pixels of the staff segments in i_{th} and j_{th} columns as illustrated in Figure 4. In our proposed method, k is set as 5.

Since the staff lines are parallel, the orientation of the staff segments within one column should be similar. The orientation of the staff line at each column can be determined by the orientations of the staff segments at that column, which is defined as below:

$$O_i = \frac{1}{p}\sum_{t=1}^{p}(Orien_t) \tag{2}$$

where O_i denotes the orientation of the staff line at $i-th$ column, $Orien_t$ denotes the t_{th} staff segment orientation on i_{th} column, and p is the total number of the staff segments in the i_{th} column. Basically, the orientation of staff line at i_{th} column is calculated by averaging the orientations of the staff segments so that the result is stable and accurate, without affecting by some outliers.

However, in some columns, the orientation of the staff line cannot be calculated due to lack of staff segments in those columns. After the previous steps, we have a list of staff line orientations $\{O_1, O_2, \ldots, O_m\}$, and those missing orientations can be interpolated using the existing list. The final estimated staff line of Figure 3(a) can be shown in Figure 3(c). The overall algorithm is summarized in Algorithm 1. Consider that one musical document usually have more than ten staff lines, the threshold T in Step 4 of Algorithm 1 is set as 10 to avoid calculating the orientation on those columns without enough staff segments.

Algorithm 1 Modeling the Staff Line Shape

Require: Initial Staff Image $InitialStaff$
Ensure: The Estimated Staff Line Model $\{O_1, O_2, \ldots, O_n\}$

```
1:  Create an empty list List_O to store the estimate staff line orientation
2:  for Each Column i = 1 to n in InitialStaff do
3:      Find out all the connected components in column i
4:      if The number of connected components is smaller than threshold T then
5:          Skip this column and go to next one.
6:      else
7:          for Each staff segments in column i do
8:              Calculate the staff segment orientation Orien using Equation 1
9:          end for
10:         Obtain the staff line orientation O_i at column i using Equation 2
11:         Add O_i to the List List_O
12:     end if
13: end for
14: Generate the staff line model {O_1, O_2, ..., O_n} via List_O using interpolation
```

C. Staff Line Detection

After the staff line shape is estimated, the musical document image pixels can be clustered as follows:

- For each pixel in the image, if it is not clustered into any groups yet, label it as a new group.
- The pixel in the new group and the estimated staff line shape will determine a curve. The curve crosses the examining pixel, and has the same orientation at each column as the staff line shape model.
- All the pixels on this same curve are clustered into one group.

All the image pixels are divided into several groups based on the staff line shape. Each group determines a curve with the same shape of staff line. If we count the white pixels(musical staff pixels) of the initial staff image of each group, the group labels that corresponding the staff line location will have large value, which is illustrated in Figure 5.

To make the staff line detection more accurate, the uniform space height of staff lines is also taken into account. Since the staff lines exist in groups and are parallel, the staff line positions have interval around $staff\_space$. Only if there exists staff line candidates at position $p - staff\_space \pm \delta$ or position $p + staff\_space \pm \delta$, a staff line candidate position p can be detected as a staff line position. δ is a user select parameter that allows some variation. Empirically it is set as $staff\_height$, and the staff line detection result is shown in Figure 3(d), where all the staff line locations are detected correctly.

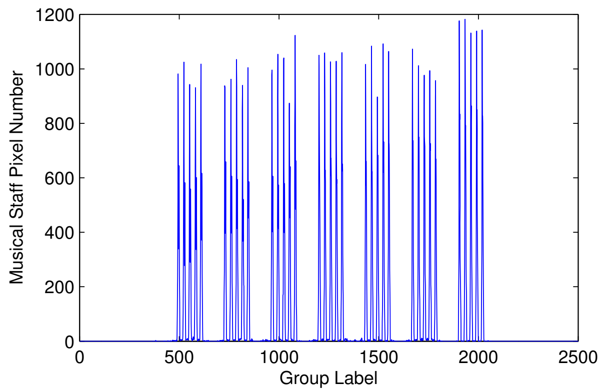


Figure 5. The histogram of musical staff pixel number on each group

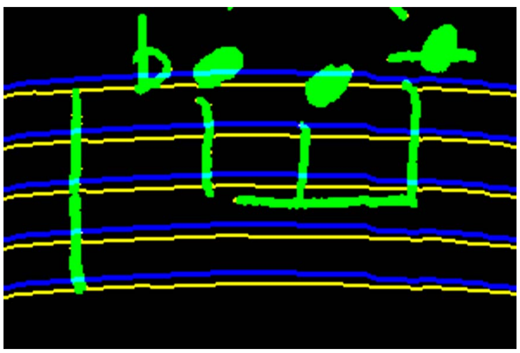


Figure 6. A fragment of Figure 3(d). The yellow lines denote the real staff lines, and the blue lines denote the estimated staff lines.

D. Staff Line Removal

Removing those staff pixels on the detected staff line location is straightforward. There are only two issues that need to be taken care.

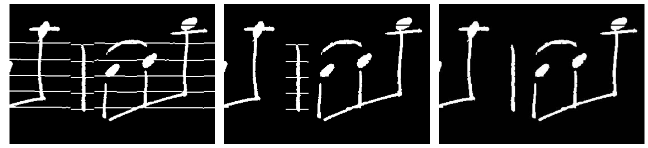
First, the staff lines are intersected with musical notes, so the staff segments on the estimated staff line location are classified as follows:

$$S = \begin{cases} \text{Staff Line} & len < T \\ \text{Musical Note} & \text{otherwise} \end{cases} \tag{3}$$

where S denotes a staff segment on one column, len denotes the vertical length of the examining staff segment S and T is a user defined threshold which is set as $MIN\{2 \times staff\_height, staff\_height + staff\_space\}$. Those staff segments with larger vertical length are simply classified as musical score and kept on the final result.

Second, the estimated staff line locations do not fully match the real staff lines on the musical document. There might be some errors during the model fitting stage, so some small gaps exists between the estimated staff line and real staff line, as illustrated in Figure 6.

So in those columns where the corresponding pixels on the estimated staff line locations are black, we will look up the nearest staff segments up and down within a local window on that column. If the staff segments can be found, it will be classified using the criteria defined in Equation 3. The local window here is set $staff\_space$. The final result is obtained as illustrated in Figure 3(e).



(a) Input Image (b) Line Tracking result (c) Our result

Figure 7. A fragment of musical document deformed by Typeset-emulation and corresponding results obtained by Line tracking method and our proposed method, respectively.

Table I
PERFORMANCE OF OUR METHOD ON ICDAR 2011 COMPETITION TRAINING DATASET

Deformation Type	Error Rate(%)
Ideal	1.33
Curvature	1.43
Interrupted	1.02
Kanungo	2.84
Rotation	1.65
Line Thickness-variation-v1	3.62
Line Thickness-variation-v2	2.89
Staff Line-y-variation-v1	4.58
Staff Line-y-variation-v2	3.64
White speckles	1.37
Typeset-emulation	2.09
Average	2.41

III. EXPERIMENTS AND DISCUSSION

The proposed method is tested using the dataset of the musical staff removal competition held under ICDAR 2011 [1]. The competition dataset consists of 1,000 handwritten music score images, and these 1000 musical documents are distorted using 11 deformation models. Each deformed model generates 1000 images. In total, this dataset consists of 12000 images, which are divided into 6000 training data and 6000 testing data. However, our method rejects those images with large staff line thickness, so all the experiments are conducted over 5500 images in training and 5500 images in testing dataset. The performance of our algorithm is evaluated using a pixel based metric, where error rate are calculated as described in [2].

Table I shows the quantitative measurement of our method on the competition training dataset. Our proposed method achieves very good accuracy rate on each kinds of deformed musical documents. Table II compares our proposed method with the winning algorithm ISI01-HA method [1] of the ICDAR 2011 competition and line tracking method described in [2] on the competition testing dataset. The

[1] http://dag.cvc.uab.es/cvcmuscima/competition/

Table II
THE ERROR RATE(%) OF OUR METHOD, ISI01-HA METHOD AND LINE TRACKING METHOD ON ICDAR 2011 COMPETITION TESTING DATASET

Methods	ISI01-HA	Line Tracking	Our Method
Error Rate	1.50	2.08(rejected 1512 images)	1.95

(a) Musical Document

(b) Removal Result

Figure 8. A fragment of deformed musical document with large staff line thickness.

error rates of our proposed method and ISI01-HA method are calculated without considering those images with large staff line thickness. Line tracking method further rejects the images deformed by curvature, interruption and rotation. Our proposed method achieved slight higher error rate compared with ISI01-HA method, but performs a little better than the Line Tracking method. However, the line tracking method rejects a few kinds of deformed musical documents [1], which can be handled using our proposed method.

In some kinds of deformations where the distortions greatly affect the staff line shape, such as curvature and rotation, our proposed method achieves the best error rate, which can be explained as the use of the staff line model. Our proposed method incorporates global information and local information together to accurately estimate the deformed staff shape model. Figure 7 shows one image fragment example taken from the competition training dataset, and the resultant images generated by our proposed method and line tracking method. Compared with the resultant image produced by line tracking method, our result has a better visual quality that retains all the musical score information without inducing noise.

The proposed staff detection and removal method is robust and effective on different kinds of deformed musical documents, but it still has some limitations. First, the estimated staff shape may not exactly match the staff line due to interpolation error. Although it can be solved by looking up staff segments within a local window as described in Section II-D, a better staff line shape model by incorporating local information can detect the staff line locations more efficiently. Second, the staff removal criteria only consider the staff segments height, it will fail to extract the musical notes when the staff line thickness is large, which is shown in Figure 8. We will focus on these two issues in our future study.

IV. CONCLUSION

This paper presents a novel music staff detection and removal technique for musical documents. The proposed method makes use of the global information of the musical document and models the staff line shape. Given a binary musical document, the proposed technique first estimates the $staff\_height$ and $staff\_space$, and constructs the initial staff line image. Then the staff line shape is modeled using the initial staff line image. After that, the estimated staff line shape is used to detect the musical staff location. Finally, the musical-score-only image is obtained by removing those detected musical staffs. The proposed technique has been tested on the latest musical staff removal competition dataset, and experimental results show the robustness and effectiveness of our proposed method on different kinds of deformed musical documents.

ACKNOWLEDGMENT

We thank Prof. Christoph Dalitz for providing his resultant images and testing datasets to us, and kindly helping us use the Gamera [2] and MusicStave [3] systems.

REFERENCES

[1] A. Fornes, A. Dutta, A. Gordo, and J. Llados, "The ICDAR 2011 music scores competition: Staff removal and writer identification," *ICDAR*, pp. 1511–1515, 2011.

[2] C. Dalitz, M. Droettboom, B. Pranzas, and I. Fujinaga, "A comparative study of staff removal algorithms," *IEEE Tran. on PAMI*, vol. 30, no. 5, pp. 753–766, May 2008.

[3] D. Blostein and H. S. Baird, "A critical survey of music image analysis," in *Structured Document Image Analysis*, Baird, Bunke, and Yamamoto(Eds.), Eds. Springer, 1992, pp. 405–434.

[4] I. Fujinaga, "Staff detection and removal," in *Visual Perception of Music Notation: On-Line and Off-Line Recognition*, S. George, Ed. Idea Group Inc., 2004, pp. 1–39.

[5] D. Bainbridge and T. C. Bell, "Dealing with superimposed objects in optical music recognition," *International Conference on Image Processing and its Applications*, pp. 756–760, 1997.

[6] R. Randriamahefa, J. P. Cocquerez, F. Pepin, and S. Philipp, "Printed music recognition," *ICDAR*, pp. 898–901, 1993.

[7] A. Dutta, U. Pal, A. Fornes, and J. Llados, "An efficient staff removal approach from printed musical documents," *ICPR*, pp. 1965–1968, 2010.

[8] H. Miyao, "Stave extraction for printed music scores," in *Intelligent Data Engineering and Automated Learning-IDEAL 2002*, H. Yin, N. Allinson, R. Freeman, J. Keane, and S. Hubbard, Eds. Springer, 2002, pp. 621–634.

[9] J. dos Santos Cardoso, A. Capela, A. Rebelo, C. Guedes, and J. P. da Costa, "Staff detection with stable paths," *IEEE Tran. on PAMI*, vol. 31, pp. 1134–1139, June 2009.

[2] The Gamera Project Homepage. http://gamera.sourceforge.net

[3] Staff Line Removal Toolkit. http://music-staves.sourceforge.net