Multimed Tools Appl
DOI 10.1007/s11042-017-5256-y

A music symbols recognition method using pattern matching along with integrated projection and morphological operation techniques

Mahmood Sotoodeh[1,2] · **Farshad Tajeripour**[1] · **Sadegh Teimori**[1] · **Kirk Jorgensen**

Received: 23 November 2016 / Revised: 20 September 2017 / Accepted: 25 September 2017
© Springer Science+Business Media, LLC 2017

Abstract Optical Music Recognition (OMR) can be divided into three main phases: (i) staff line detection and removal. The goal of this phase is to detect and to remove staff lines from sheet music images. (ii) music symbol detection and segmentation. The propose of this phase is to detect the remaining musical symbols such as single symbols and group symbols, then segment the group symbols to single or primitive symbols after removing staff lines. (iii) musical symbols recognition. In this phase, recognition of musical symbols is the main objective. The method presented in this paper, covers all three phases. One advantage of the first phase of the proposed method is that it is robust to staff lines rotation and staff lines which have curvature in sheet music images. Moreover, the staff lines are removed accurately and quickly and also fewer details of the musical symbols are omitted. The proposed method in the first phase focuses on the hand-written documents databases which have been introduced in the CVC-MUSCIMA and ICDAR 2013. It has the lowest error rate among well-known methods and outperforms the state of the art in CVC-MUSCIMA database. In ICDAR 2013, the specificity measure of this method is 99.71% which is the highest specificity among available methods. Also, in terms of accuracy, recall rate and f-measure is only slightly less than the best method. Therefor our method is comparable favorably to the existing methods. In

✉ Mahmood Sotoodeh
mahmood.sotoodeh@aggiemail.usu.edu

Farshad Tajeripour
tajeri@shirazu.ac.ir

Sadegh Teimori
sta.teimori@gmail.com

Kirk Jorgensen
kirk.jorgensen@aggiemail.usu.edu

[1] Department of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran

[2] Department of Computer Science, Utah State University, Logan, UT 84322-4205, USA

Published online: 16 November 2017

the second phase, the symbols are divided into two categories, single and group. In the recognition phase, we use a pattern matching method to identify single symbols. For recognizing group symbols, a hierarchical method is proposed. The proposed method in the third phase has several advantages over the previous methods. It is quite robust to skewness of musical group symbols. Furthermore, it provides high accuracy in recognition of the symbols.

Keywords Optical music recognition · Integrated projection · Morphological operation · Pattern matching · Group symbols · Single symbols · Staff lines

1 Introduction

Recognizing the musical alphabet is essential and preliminary to read and perform music. There are seven musical symbols to indicate pitch which are acknowledged worldwide by two types:

A. Syllable: used in countries like France, Italy and Iran is derived from a medieval religious poem written by a French priest, Guido d' Arezzo (tenth century).

Do Re Mi Fa Sol La Ti

B. Alphabetical: used in Germany, Austria, England and America, and is as follows:

C D E F G A B.

The musical staff consists of five parallel, horizontal lines. Musical symbols are placed on and between staff lines. Staff lines are numbered bottom-up, from 1 to 5. As the symbol goes upward it has a higher pitch, and as it goes downward it becomes lower in pitch. There are directive symbols called "clefs" (e.g. sol clef) which change the baseline for enumeration of notes (see Fig. 1). The number of music symbols is abundant but they are a finite set. Some important music symbols are shown in Table 1.

The use of computers in recent decades has had a dramatic impact on the advancement of different sciences. Today computers are used impressively in music. Music specialized software provides various facilities for those who like and work with music. Remarkable collections of musical symbols (e.g., sheet music image) are available in different sources, such as internet. Sheet music images are created by scanning the musical book's pages, or by typing in specific programs. Then they are saved as images. However, for blind and visually impaired people who enjoy music, learning music from printed books or note sheets is impossible. Even learning the symbols written in Braille is too difficult for them. So, there is a need to provide musical symbols in a practical way for the blind and low-vision people. This requires presenting image processing algorithms to extract symbols sequences from their images. After extracting symbols sequences (music sentences), it is possible to synthesize them into the human voice using other software to perform solfège[1] of the musical notes. Such a system is not only applicable for a blind person, but it is also efficient for children, beginners, composers and musicians.

[1] https://en.wikipedia.org/wiki/Solfege

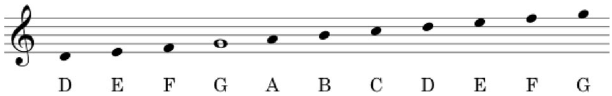
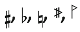
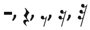


Fig. 1 Position of symbols on staff lines

Extracting Musical symbols from images is performed by a system similar to the Optical Character Recognition (OCR) and is called Optical Music Recognition (OMR). The aim of OMR is accurate recognition of the sequence of musical symbols of a song. After extraction, symbols can be saved in an appropriate data file format such as standard IEEE XML format [2] or Music Instrument Digital Interface (MIDI) which is a readable format for electronic musical instruments. Furthermore, the extracted note sequences, like language sentences, can be processed, indexed and searched.

Table 1 Definitions of some important music symbols

Symbols	Definition
	Beam: A thick line frequently used to connect multiple consecutive single notes with flags (e.g. eighth notes).
	Single Note with flags (Eighth note, Sixteenth Note, Thirty second Note, Sixty-fourth Note and others with more flags)
	Clef: Bass, Alto and Treble, the first symbols that are seen in the beginning of the staff line and that change the order of notes.
	Sharp, Flat, Natural, Sori, Koron are placed after clef and before the Notes to raise and lower the sounding pitch.
	Half, quarter, eighth, sixteenth and thirty-second rest indicate the exact duration of silence in the music.
	Slurs and Ties: Ties and slurs may be confusing because they are both represented by a curved line. However, the function of a tie is very different from a slur. A tie is a curved line that connects two notes of the same pitch; the second note is not played but its value is added to the first note. On the other hand, a slur connects two or more notes that have the same or different pitch.
	Note head, Quarter, Half, Note (a main part of note. A note having one-fourth the time value of a whole note. Also called crotchet. A note having one half the value of a whole note. A note having the value of four beats in common time.
	Volta brackets – or “time bars” – are horizontal brackets labeled with numbers or letters that are used when a repeated passage will have two or more different endings.

First of all, in this research, we introduced projection and morphological operation technique to detect and remove staff lines in sheet music images. In this step, we used thickness and space between two staff lines as two important features to create structural elements for morphological technique. After that, we expressed and analyzed the disadvantages and advantages of each method. We found out that advantages of each method resolved the disadvantages of each other. Finally, we decided to combine them in order to retain the advantages of each method and eliminate the disadvantages of them almost completely. Therefore, we could obtain a good result in this phase with this trick.

For detecting musical symbols, we proposed a method in order to separate musical symbols in two classes; single and group symbols. The single symbols can have appeared alone in sheet music images. In addition, some of them like notes are able to combine to each other by connecting to beam symbols and create new group symbols. Therefore, a group symbols is a set of single symbols like notes which connect to gather by beam symbols. For separating the group symbols from single symbols, first we performed connected component technique to obtain a bounding box for each symbol in sheet music images. After that we calculated three important features for each component in the sheet music image. These features were width, height and area. We needed a good threshold to compare features for separating the group symbols from the single symbols. We empirically found out the best value for threshold was calculated as eighteen times the power of two of the distance value between two horizontal staff lines in sheet music images. Therefore, for each component which its area was bigger than this threshold it was considered as candidate group symbols. Some musical symbols were not group symbols but they had this feature and were selected for group symbols, for example clefs. We applied width to height ratio as a feature to group candidate image which was obtained from previous step. If this feature for any candidate group symbols was less than 1 it was known as a single symbol and was removed from group candidates image. Finally, we have single symbols image(SS) and group symbols image(GS).

We applied two methods for recognizing musical symbols. We used pattern matching for single musical symbols image and used a simple hierarchical processing to classify group symbols. Finding suitable pattern for any symbol music class was a challenge because we wanted to have a pattern matching approach which was robust to rotation of single music symbols. For solving this problem, first we computed the average of each class and used them as representative patterns. After that, for obtaining a robust pattern matching approach, we rotated each representative of each musical symbol class by ten degrees and we did this thirty-six times. Then, the similarities between them and each component of single symbol image(SS) were computed. Next, the similarity value of each component were compared with a threshold. If it was less than the value of the threshold which has been subtracted from one, then this object was chosen as the best pattern. Finally, we used the best pattern to slide on single symbol image(SS) to find the symbols with the highest match to the best pattern. Then, they were labeled as same as the label of the best pattern.

There was a problem in group symbols. They had a skewness which reduced the accuracy during identifying musical symbols. For this reason, we proposed a transformation model in order to remove the skewness of the group symbols. After the skewness correction, we used a hierarchical method to recognize symbols in group symbols. First, we perform opening filter to remove stems. Stems are vertical lines which connect to the note heads. The goal of this step was to separate note heads from beams. We chose a horizontal structure element whose size was twice the thickness of the staff line. As was mentioned before, two important features in sheet music were thickness of staff line and vertical distance between two staff lines. After removing stems, sometimes there were a small meaningless objects remain on sheet music image, we eliminate them by opening filter with vertical structure element whose size was twice the thickness of the staff line. After that, in order to separate beam symbols and note head symbols, we applied connected component technique. For each component, we defined some conditions and thresholds to separate note heads from the beams. If defined conditions would occur for each component, then we labeled it as beam symbol otherwise it was labeled as note head. For identifying the type of each note, the number of beams which were located below or up from the center of each note head were counted.

The aim of this study is to develop an efficient algorithm for extraction and detection of the musical note sequence. The output can then be fed to a software which solfège the musical symbols (e.g., Do, Re, Me, Fa, sol, La, Ti) to help blind or visually impaired people, children, and beginners. Extracted symbols consume much lower capacity during storage compared to images. These symbols can also be used in other researches on intelligent algorithms and data mining. This paper is organized as follows: a review of recent researches is presented in the next section. In the third section, the proposed methods for each phase are given, and in section 4 the results are analyzed and compared with other methods. Finally, conclusions and suggestions for future work are presented in section 5.

2 Background review

Investigation in OMR started in the 1960 by Pruslin [30] and Prerau [29]. In parallel with the development of technology and computers, several commercial software applications for OMR have entered the market. Each has its own advantages and disadvantages. Moreover, although these softwares are available, there are still many challenges, especially for hand-written musical scores. Application of OMR on portable cameras is also currently under investigation [21, 39, 46].

2.1 Detection and removal of staff lines

A musical page consists of staff lines and musical symbols. The two important phases of OMR system are detection and removal of staff lines followed by recognition of musical symbols. In the following sections we investigate some of the existing algorithms for these two phases.

Although, there are algorithms that work without removing the staff lines [3, 5, 23, 31, 32, 42] the presence of staff lines complicate the analysis and detection of musical symbols, since it combines different parts into complex grouped connected components. So it is important to detect and eliminate the stuff lines.

David Bainbridge et al. [1] used the combination of horizontal image and histogram technique to detect staff lines. Fujinaga [16] utilized image processing techniques, such as connected components and run length coding to efficiently extract two features of staff lines, line thickness and vertical distance between lines. A problem with this method is that some components which were not part of staff lines but had the same thickness, were also removed. Additionally, symbols such as ties and slurs are repeatedly considered as staff lines. Since these methods are based on connected components analysis, they are very slow. Chen Genfang et al. [17] used Hough Transform and morphological operations to detect lines, but the time complexity is its drawback.

Tajeripour and Sotoodeh [43] used the vertical projection and run length coding techniques for detecting staff lines. Then, to remove the effect of noise and reconstruct the solid staff lines, erosion and dilation operations were used. Identifying volta brackets (a symbol whose thickness is equal to the thickness of staff lines) was a positive point of this method. The disadvantage of this method is that it assumes that staff lines are straight. Therefore, with some rotation and curvature in staff lines, accuracy decreases. Slurs and ties are two important symbols in sheet musics. Their thickness is approximately identical with that of staff lines. Sotoodeh and Tajeripour [40] proposed first and second order derivative and connected component analysis to detect these symbols and discriminate them from staff lines.

In another study, Miayo et al. [23] considered long objects in sheet music images which consist of a series of points that were connected horizontally and had a stable equal vertical distance, as stuff lines. They used dynamic programming for detecting staff lines. Dalitz et al. [10] utilized the skeletonization method to extract staff lines information, and then remove them. Cardoso et al. [37] regarded musical images as a graph and staff lines as connected paths, starting from the left margin and ending to the right margin. Paths which were the shortest and also which have enough black pixels are chosen as staff lines. Ng et al. [27] also used the image skeleton in their analysis. The advantage of this model was that analysis of the skeleton was easier and more straight forward.

Dutta et al. [11] utilized segments obtained from run-block coding and based on properties such as height, direction and distance, they discriminated between staff line parts and other segments. Bolan Su et al. [41] removed staff lines by using thickness features and the distance between two consecutive lines and then used the images with estimated staff lines to create a model. This model helped to estimate staff lines more accurately. Timofte and Luc Van Gool [44] used binary image techniques and dynamic programming to detect staff lines and removed the noise in images. Nhat, Vo Quang and Guee Sang Lee [28] applied two features of staff lines, thickness and the distance between consecutive lines to estimate candidate staff line segments, then staff lines were extracted by selecting from this group of candidates. Géraud et al. [18] used a morphological based method to detect and remove staff lines in sheet music images. This method has been introduced as the LORD method in recently published papers. In addition, it was accepted in the latest conference as the best method in the ICDIR 2013 competition for removing staff lines. Montagner et al. [24] proposed a machine learning based method to remove staff lines in sheet music images which they called FS. This method trained suitable windows for morphological operations and then used them to eliminate staff lines effectively. Manuel was another learning based method which was proposed by Montagner et al. [38] which was based on Hirata [20]. They trained image operators and utilized them to detect and remove staff lines. In [25] a learning algorithm was introduced. It was based on an optimization method and named NILC. This method followed the Manual method which was presented in [38]. Two important stages of this method were to determine the window operators and to minimize operator errors simultaneously. In [26], Kernel SVM has been used to estimate window operator learning to detect and remove staff lines in sheet music images. Good generalization and nonlinear decision surfaces estimation were two important advantages of this method. This method is known as the Kernel Approximation(KA). Calvo-Zaragoza et al. [6] used classifiers to separate staff lines from non-staff lines and finally, removed them. Subsequently they developed this method for detecting and removing staff lines in gray scale images instead of binary images directly [7]. In this research, we propose a simple but accurate method for staff line removal in handwritten pages. It is both accurate and enjoys high speed. Compared to the existing algorithms it can remove staff lines with higher accuracy.

2.2 Detection of musical symbols

In this section, we investigate the algorithm in the second and third phases of OMR. The aim of this part is to determine and classify musical symbols. Mahoney et al. [22] built a set of candidates to one or more symbol types and then used descriptors to select the matching candidates. Carter [8] used LAG (line adjacency graph) techniques to extract symbols. The symbols were classified based on bounding box size. Randriamahefa et al. [33] introduced a structural method based on graph. In this method, for each symbol a graph was selected, and in the following phases, each graph was separated by Region growing and thinning techniques. Cousnon et al. [9] also introduced a method which was based on extraction grammar from musical knowledge data (See Fig. 2).

Toyama et al. [45] extracted a set of rules with pixels tracking to create a model for recognizing musical symbols. Rossant and Bloch [36] introduced another method which worked based on a fuzzy model and a model of pattern matching. The advantage of this method was its robustness with the changes that happen in symbols.

Rebelo et al. [34] carried out an investigation on four classifiers, namely Hidden Markov Models, Nearest Neighbor (KNN), Neural Networks (NNs), and Support Vector Machines (SVMs). The best accuracy among these classifiers was Support Vector Machine. In [35] a set of musical rules were considered as basic knowledge to extract musical symbols. For example, in a musical page, clefs symbols appear at the left margin of staff lines. This method checked to see whether there was a clef or not, if a clef existed, in the next phase it checked whether there was a key signature symbol or not. Indeed, by using sequences of musical rules, the extraction of musical symbols could be performed. Zaragoza and Oncina [4] introduced an online method to recognize the hand written musical symbols which were written by stylus pens on portable devices such as tablets and iPads.

In each phase of these methods there are some strong and weak points in detecting staff lines and musical symbols which can be seen both in handwritten and printed symbols. Regarding computation, some of these methods are fast, but the accuracy in detecting lines or detecting musical symbols is low. On the other hand, some methods have low computing speed, but their accuracy in detection in the first and second phases is high. The proposed method in this research is robust toward changes in lines because it is not dependent on the rotation of staff lines, and in the second phase, it is also robust to deviation of the symbols positions and their rotation.

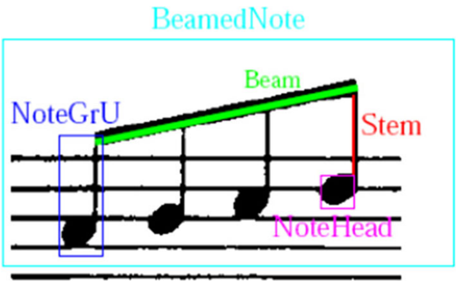


Fig. 2 Determination of musical symbols by grammar extraction [9]

3 The proposed method

In this research, we extract two important features of staff lines which are used in all phases. They are computed based on [43]. One of them is thickness of staff lines and another one is the vertical distance between two lines. They are shown in Fig. 3.

In this paper, first the proposed algorithm for removal of lines will be explained, and then the detection and recognition of musical symbols will be discussed.

3.1 Detection and removal of staff lines

The method used for removing staff lines is a combination of horizontal projection [16] and also morphological operations. These two methods are explained below:

3.1.1 Horizontal projection

In this method, we compute and use the length of the staff lines as an important feature to detect the staff lines. It is common for the length of lines to be near 70% of sheet music width and to have the largest length in the set of musical symbols, therefor the horizontal projection value relating to staff lines has the maximum value. If, for example the original image is I_0, then the horizontal projection on it can be computed by eq. 1.
for each row r in image I_0

$$\mathrm{HP(r)} = \sum_{c=1}^{Columns} \mathrm{I_0(r,c)} \tag{1}$$

In eq. 1, HP(r) is projection along the rows, I_0 is the image of sheet music, c and r indicate the index of column and row of image pixels respectively. After obtaining horizontal projection by applying a suitable threshold which is 65% of sheet music width in eq. 2, the location of staff lines can be found.
for i = 1 to length of the HP

$$staff\ \ line(\mathrm{i}) = \begin{cases} 1 & if\ \mathrm{HP(i)} > 0.65^{*} width \\ 0 & otherwise \end{cases} \tag{2}$$

After the detection of staff lines for removing them, XOR operator is performed on the original image and the staff line image according to eq. 3. The results have been shown in Fig. 4.

$$\mathrm{I_{Sym}} = \mathrm{XOR}\ (staff\ \ line, \mathrm{I_0}) \tag{3}$$

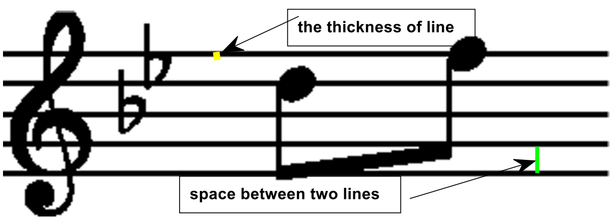


Fig. 3 Two important features of staff lines. The thickness of line and the vertical space between two lines

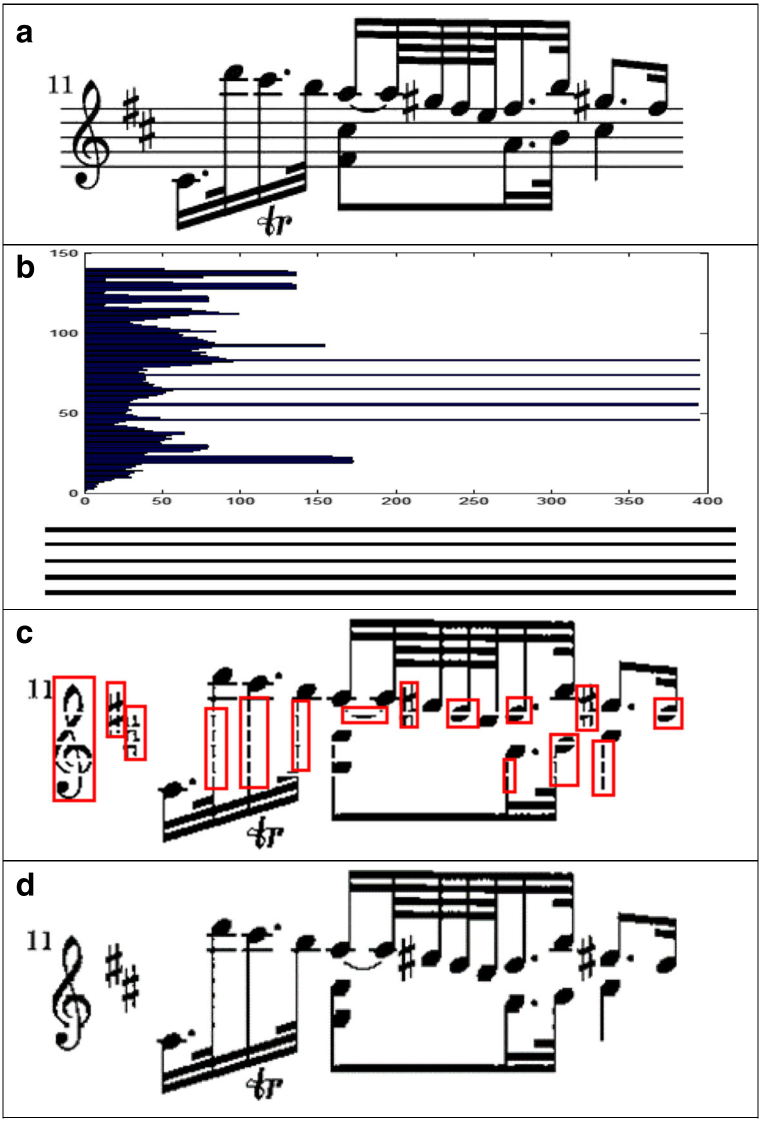


Fig. 4 Staff lines detection and removal by Horizontal projection (**a**) Original image, (**b**) Horizontal projection, (**c**) Staff lines, (**d**) Image without staff lines. In this part, the red color indicates the failures of the method, (**e**) The ground truth image

In eq. 2, for any object in sheet music image, if the obtained value of horizontal projection is greater than the 65% of sheet music width, it could be a staff line. I_{Sym} in eq. 3 is the sheet music image without staff lines. This method has it is own advantage and disadvantages. The advantage of this method is that it gives us the position of staff lines. In addition, it is very fast and simple.

The disadvantages of this method are that it segments image symbols, and also if the staff line has curvature or rotation, it does not work well, especially in the case of hand-written musical notes, the staff lines cannot be properly removed. The advantages are shown in Fig.

4d. by red color. For example, some stems, note heads, sharps, slur and clefs are fragmented after removing staff lines.

3.1.2 Staff lines removal with morphological operators

Morphological operator is a branch of biology dealing with the study of the form and structure of animals and plants. Mathematical morphology uses this term as a tool to show and describe the shape (such as boundary, skeleton, and convex hull). In addition, as a morphological filter it is used to thin and prune. Morphological operator includes four basic operations; erosion, dilation, opening, and closing [19]. In this paper the opening filter is used to remove the staff lines. Opening operation is an erosion(⊖) followed by a dilation(⊕). The erosion shrinks the foreground and therefore it is useful to eliminate noises and the staff lines from the sheet music image. Although it removes the staff lines, but simultaneously it reduces the size of the music symbols in the images. For this reason, erosion is followed by dilation using the same structural element. In this way, since the small-scale components are gone, they won't come back, but the size of remained components are increased. Suppose $I_{Original}$ and SE show the image and the structural element, respectively. The opening operation is defined as eq. 4.

$$\mathrm{I}_{Opening} = \left(\left(\mathrm{I}_{Original} \ominus \ \mathrm{SE}\right) \oplus \ \mathrm{SE}\right) \tag{4}$$

The structural element (SE) is a small matrix or vector of pixels, each with value of zero or one. The pattern of ones and zeros determines the shape of the structural element. For opening, the structural element SE is slide over the image $\mathrm{I}_{\mathrm{Original}}$ (as in 2D convolution). In erosion, a pixel in the image will be considered 1 only if all the pixels under the SE is 1, otherwise it is eroded, i.e. made to zero. As the result, the noise will be removed and the thickness of the foreground object decreases. In dilation, a pixel in the image is considered as 1 if at least one pixel under the SE is 1. So, it increases the white region in the image or size of foreground objects.

Here, in order to eliminate the staff lines, a vertical-shaped structural element with the size of $(h \times 1)$ is used for opening (Fig. 5). The size of SE (h) should not be much thicker nor thinner than the staff lines. If its size was small, the staff lines will not be removed, and if the size was large, more details will be removed in the images. Here, the size h is set to the thickness of the staff lines plus one, i.e. $(t + 1)$, which provides an appropriate size of the structural element for each sheet music image. The disadvantage of this method is that the image components with similar thick staff lines (such as slurs and ties) are removed, and the image components are fragmented (see Fig. 5).

3.1.3 Integrating two methods of horizontal projection and morphological operation

The investigation of these two methods shows that they have some weaknesses. For example, they may not only remove the staff lines, but also remove other cursory lines as well, such as volta brakets, auxiliary staff lines, slurs and ties. In addition, some parts of musical notes maybe fragmented, especially the notes that stick to staff lines. According to eq. 5, if we use OR operator between two outputs in these two methods, they can cover each other's weakness and a better result can be seen (black pixels have 1 and white pixels have 0). The results of three methods are shown in Fig. 6.

$$\mathrm{I}_{[\mathrm{Sym}-Opening]} = \mathrm{I}_{\mathrm{Sym}} OR\ \mathrm{I}_{Opening} \tag{5}$$

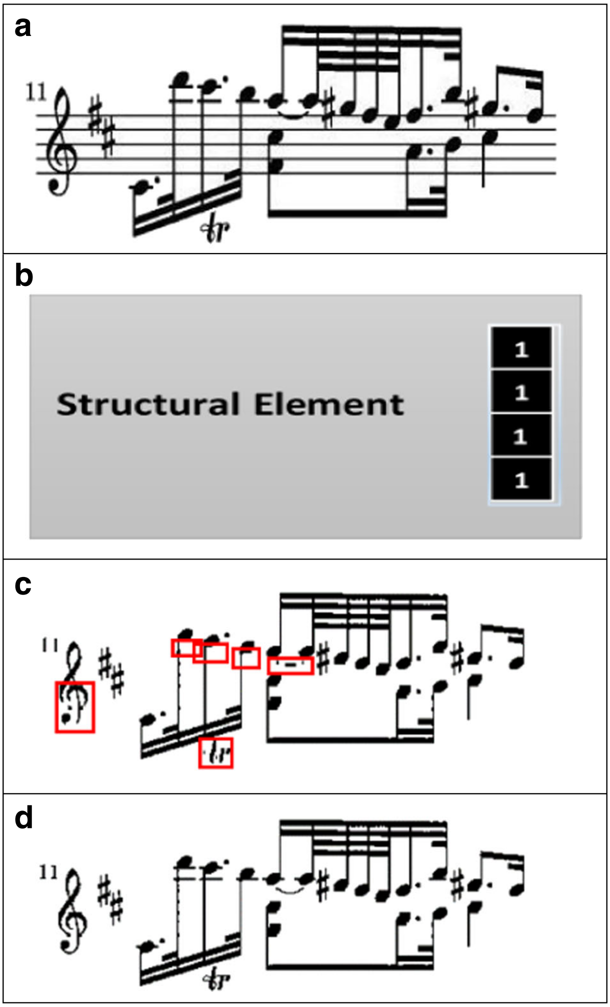


Fig. 5 Removing the staff lines by using morphological operators. **a** Original image, **b** Structuring element **c** Staff line removal using morphological operation (The red rectangles indicate the failures of this method.). **d** The ground truth image

In eq. 5, I_{sym} is the staff line removal result of horizontal projection which includes the music symbols and $I_{Opening}$ is the sheet music image without any staff lines which is obtained by morphological operation. $I_{[Sym\text{-}Opening]}$ is the image result of integrating morphological operation and horizontal projection. In Fig. 6., the result of each method has been shown. As you can see, the output image of horizontal projection method could not keep the structure of some symbols when it compares with ground truth image. This method has several failures which are indicated by red color rectangles. The result of morphological operations is better than the result of horizontal projection. However, in comparison with the ground truth image, there are some failures which had not appeared in horizontal projection such as the trill symbols. They are shown as red color rectangles. Since the images are binary we can use the binary operators on pairs of images in each step of the method. Finally, we decided to choose OR operator to correct the errors.

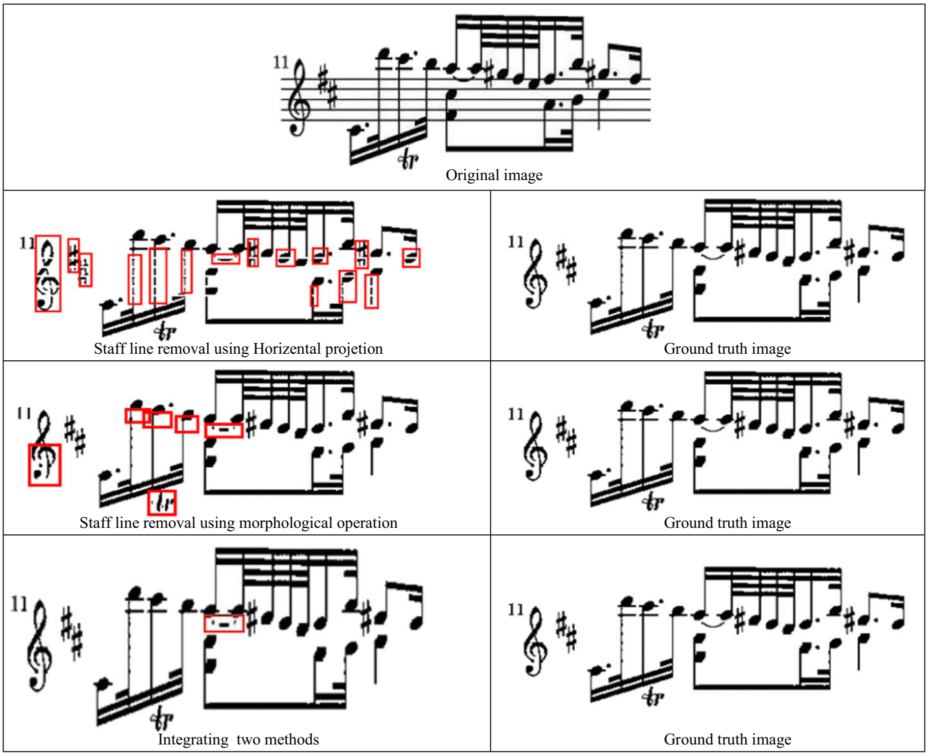


Fig. 6 Staff lines removal by integrating two methods of horizontal projection and morphological operation. The top part of figure is original image. The left column shows the result of each method and the right column indicates the ground truth or target image (The red color rectangles show the fragmented music symbols after removing staff lines)

3.2 Detection and identification of musical symbols

3.2.1 Detection symbols and classify them into two categories.

The symbols are classified as group symbols and single symbols after removing the lines. Group symbols are created from several single symbols. Fig. 7. shows the sample of single symbols and group symbols.

The simplest way to separate these two categories is to use an appropriate threshold on symbols size. First, the bounding boxes of all symbols are calculated by connected component technique. It is called CC (connected component). Then for each component, width, height and area are computed. Those components whose areas are greater than a threshold are considered as group symbol candidates. This threshold is a function of the vertical distance between two

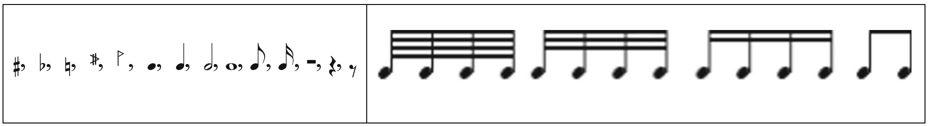



Fig. 7 some samples of single and group symbols. The left column shows the single symbols and right column shows group symbols

staff lines. It is computed as eq. 6. And then it is used in eq. 7 to separate group symbols from single symbols.

$$T = 18^{*}(\mathrm{VD\_L})^{2} \tag{6}$$

$$\textit{For each element } \mathrm{c} \textit{ in } \mathrm{CC}\begin{cases} \mathrm{c} \in \mathrm{CGS} & \textit{if } (\textit{area of } \mathrm{c}) > T \\ \mathrm{c} \in \mathrm{CSS} & \textit{Otherwise} \end{cases} \tag{7}$$

In eq. 6, T is threshold and VD_L is the vertical distance between two lines which is computed based on [43]. CGS and CSS in eq. 7 are the set of candidate group symbols and candidate single symbols receptively. If the area of each component in the image was greater than threshold T then it is placed in CGS set, otherwise it is placed to CSS set.

After using threshold, there are a few symbols in CGS which are not group symbols. Since the area of them is nearly same as the group symbols they are placed in CGS set (for example Clef symbols). For solving this problem, we use the width to height ratio for each element in GSS set to separate single symbols from CGS set and add them to CSS set as in eq. 8.

$$\textit{For each element } \mathrm{i} \textit{ in } \mathrm{CGS}\begin{cases} \mathrm{i} \in \mathrm{GS} & \textit{if } \left(2^{*}\textit{width}_{\mathrm{i}}\right)/\textit{height}_{\mathrm{i}} > 1 \\ \mathrm{i} \in \mathrm{SS} & \textit{Otherwise} \end{cases} \tag{8}$$

After this step, the CGS set is called group symbols image or GS set and CSS set is called single symbols image or SS set. The result is shown in Fig. 8.

3.2.2 Recognition of symbols in single symbol image(SS)

After classifying the symbols into two images sets, group symbols(GS) and single symbols(SS), it is time to identify them. Single symbols can be easily identified by using pattern matching. For finding a suitable pattern in an image, we use all possible windows to slide on to sheet music image. The window size is equal to the size of image patterns. For each window, a similarity between the window and the desired pattern is computed by eq. 9.

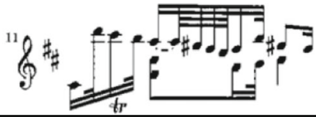
$$\textit{Similarity} = \frac{1}{\mathrm{w}^{*}\mathrm{h}} \sum_{\mathrm{x}=1}^{\mathrm{w}} \sum_{\mathrm{y}=1}^{\mathrm{h}} \mathrm{abs}(\textit{window}(\mathrm{x.y}) - \textit{pattern}(\mathrm{x.y})) \tag{9}$$

In eq. 9, x and y indicate the coordinate of points inside the windows. The w and h are width and height of the window respectively. The only challenge is finding a suitable pattern. For this, we use training data for each symbol which is introduced in [34]. Each class of particular symbol has several training symbols. For example, for a sharp symbol (#), we have 50 different training image data. For each class, we compute the average of the training image data and then we call them good representative for each class. The average of training image data for each class can be calculated by eq. 10.

for each class c in C.

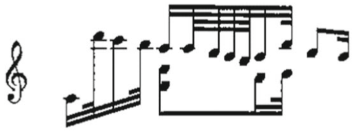
for each element i in class c

$$\begin{array}{l} \textit{Average}\{\mathrm{c}\} = \dfrac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \textit{training image data}_{\mathrm{i}}(\mathrm{x}, \mathrm{y}) \\ \quad \text{end} \\ \text{end} \end{array} \tag{10}$$

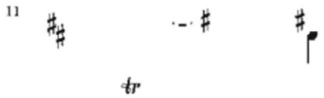


# of components	Width	Height	Area
1	4	11	44
2	0	0	0
3	4	10	40
4	22	69	1518
5	9	26	234
6	9	28	252
7	98	105	10290
8	21	19	399
9	149	66	9834
10	146	55	8030
11	1	2	2
12	8	2	16
13	1	2	2
14	9	27	243
15	10	27	270
16	45	36	1620
17	11	36	396

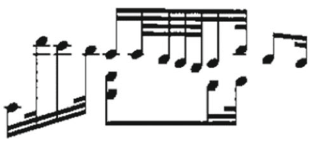
The result after using the first threshold

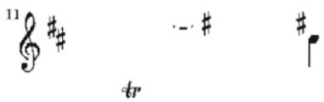


# of components	Width	Height	Area
4	22	69	1518
7	98	105	10290
9	149	66	9834
10	146	55	8030
16	45	36	1620



# of component	Width	Height	Area
1	4	11	44
2	0	0	0
3	4	10	40
5	9	26	234
6	9	28	252
8	21	19	399
11	1	2	2
12	8	2	16
13	1	2	2
14	9	27	243
15	10	27	270
17	11	36	396

The final result after using the second threshold

# of component	Width	Height	Area
7	98	105	10290
9	149	66	9834
10	146	55	8030
16	45	36	1620



# of component	Width	Height	Area
4	22	69	1518
1	4	11	44
2	0	0	0
3	4	10	40
5	9	26	234
6	9	28	252
8	21	19	399
11	1	2	2
12	8	2	16
13	1	2	2
14	9	27	243
15	10	27	270
17	11	36	396

◀ **Fig. 8** Separating symbols into two sets. The top part of the figure shows the sheet music image after removing staff lines, the left columns and right columns indicate the steps of obtaining the group symbols image(GS) and the single symbols image(SS) by using defined threshold on feature matrix respectively

In eq. 10, N is the number of training images data in each class, x and y are the coordinates of points inside the window of training image data, C is an array which keeps all the existing classes. The letter c indicates the specific class and Average is an array which holds the average of each class.

After obtaining the average of each class, for having a pattern matching approach which can be invariant to rotation of single symbols, we rotate the average of each class 36 times. Each time the angle of rotation increases by ten degrees. Therefore, we have 36 representatives for each element in Average array. The rotations of average of some classes are shown in Fig. 9.

The all of the steps for finding the best pattern to recognize the single symbols should be as follows:

1. The process for recognizing of single symbols:

   1.1. The average training image data which was computed in eq. 10 should be resized as single symbols.
   1.2. Rotate each element in Average array by 10 degrees for 36 times.
   1.3. The connected component technique is used on SS image (see Fig. 10.).
   1.4. For each component in SS image

      1.1.4. and for each element in array which was rotated in step 1.2, the percentage of similarity between it and all components in SS image is computed by eq. (9).
      1.2.4. If the percentage of similarity is more than (75%), then it is considered as a candidate desired pattern which it and its similarity are added to an array.

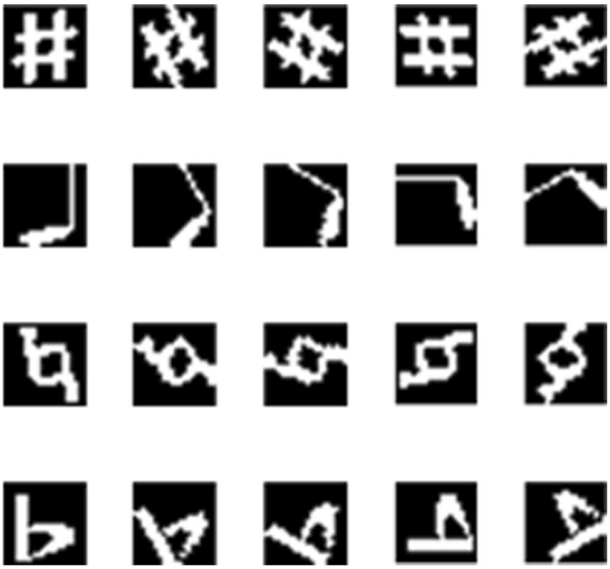


Fig. 9 Rotating the average of each class before finding the best desired pattern

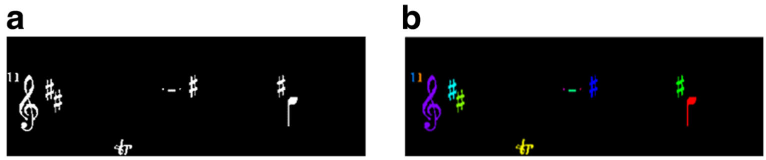


Fig. 10 Connected component technique on SS image. **a** SS image. **b** The result of connected component on SS image

1.5. In an array where the similarities are kept, the component that has the highest similarity is considered as the desired pattern.
1.6. We use the desired pattern as a window and slide it on SS image to find the location of objects which have the highest similarity to desired pattern and then label them as same as the label of desired pattern.
1.7. If the pattern could not be found for specific symbols, it means that a special symbol could not exist in the sheet music image. The result of pattern matching algorithm for identifying the music single symbols is shown in Fig. 11.

3.2.3 Analyzing and identifying the element of group symbols image(GS)

Proposed a transformation model to preprocess the group symbols before analyzing The group symbols usually have some kind of skewness; this skewness decreases the accuracy of the research method. This error is due to considering the length and width of the rectangle of every object as length and width of the object; for example, if a long object is placed horizontally, the length and width of the rectangle will be the same as the object, but if it has some kinds of skewness the length and width of the object and rectangle will not be equal. For removing this skewness, we propose a simple transformation model that completely removes this error. We apply this transformation on all pixels of image object and replace them until the corrected image of the object is created. In order to use the transformation model on every object in GS image, we need the bounding box of them. Every bounding box is displayed by four values: [x, y, width, height] which are computed by eq. 11.

$$\begin{aligned} \mathrm{X} &= \underset{x}{argmax}\left\{index\left(\mathrm{v_p}>0\right)\right\} \\ \mathrm{Y} &= \underset{y}{argmax}\left\{index\left(\mathrm{h_p}>0\right)\right\} \\ width &= \underset{x}{argmax}\left\{index\left(\mathrm{v_p}>0\right)\right\}-\mathrm{X} \\ height &= \underset{y}{argmax}\left\{index\left(\mathrm{v_p}>0\right)\right\}-\mathrm{Y} \end{aligned} \tag{11}$$

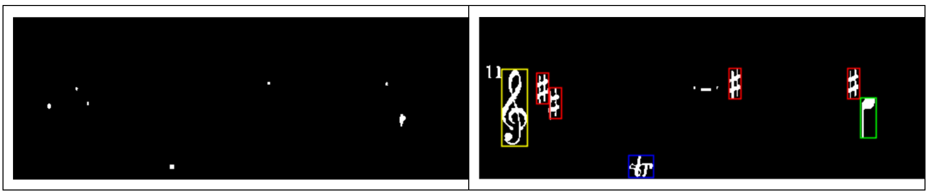


Fig. 11 The identifying single music symbols in sheet music image. In the left image, the white points indicate the locations which have the highest similarity to its desired patterns. In the right image, the colored rectangles indicate the label of each class which has been selected for any symbols in the SS image

In eq. 11, h_p is horizontal projection and v_p is vertical projection of desired symbols in GS image. After obtaining all the bounding boxes for each object in GS image, the skewness of each object is removed by the transformation model which is defined by eq. 12 and is shown in Fig. 12.

$$
\begin{aligned}
& dX = Xp{-}Xc \\
& \begin{bmatrix} Xn \\ Yn \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \tan(\theta) & 1 \end{bmatrix}^{*} \begin{bmatrix} Xp \\ Yp \end{bmatrix} + \begin{bmatrix} 0 \\ -Xc^{*}\tan(\theta) \end{bmatrix} \\
& \quad = \begin{bmatrix} Xp \\ Xp^{*}\tan(\theta) + Yp \end{bmatrix} + \begin{bmatrix} 0 \\ -Xc^{*}\tan(\theta) \end{bmatrix} \\
& \quad = \begin{bmatrix} Xp \\ (Xp{-}Xc)^{*}\tan(\theta) + Yp \end{bmatrix}
\end{aligned}
\tag{12}
$$

In eq. 12, dX is the horizontal distance of every pixel to the center of image object, and dY is the vertical distance of every pixel to the center of image object. The angle θ is the rotation angle of beam symbol relative to horizontal axis. Xc and Yc are the central pixel coordination of each image object, Xp and Yp are the coordination of each pixel from object, Xn and Yn are the new coordination of the pixels of the object. The group symbols only have skewness in y direction. Therefore, the value of Xp does not change($Xn = Xp$) but the value of Yp is replaced by the new value($Yn = (Xp - Xc)^{*} \tan(\theta) + Yp$). In eq. 12, dy is not used because the objects are skewed only in y direction and it is not needed to change the value of Xp. If we want to use dy in the eq. 12 the Xp value will change.

The angle of beam relative to the horizontal axis is calculated as follows: first, we use morphological operation to remove all the vertical lines such as stems for each group symbols in GS image according to eq. 13.

For each group g in GS image

$$I_{beam(g)} = GS_{(g)} \ominus SE \tag{13}$$

In eq. 13, GS is an image which includes the group symbols. The element g indicates the index of each group symbols in GS. SE is the structural element which is used to remove vertical lines in GS images. The type of structural element is rectangle whose size is [1*(2*thickness of staff lines + 1)]. I_{beam} is the image without vertical lines. The result of this step is shown in Fig. 13.

After removing the stems, for all the beams in I_{beam}, the morphological closing is utilized to convert the all the beams to a unique line according to eq. 14. (see Fig. 14.)

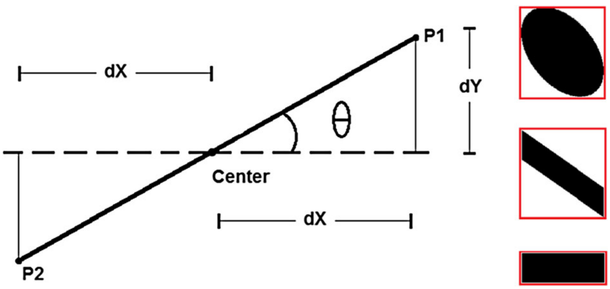


Fig. 12 the transformation model for removing skewness of objects in image

Fig. 13 Removing the vertical line (stem) by using morphological erosion

For each element i in I_{beam}

$$I_{BL}(i) = ((I_{beam}(i) \oplus SE) \ominus SE) \tag{14}$$

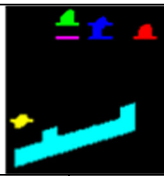
In eq. 14, I_{beam} is an image which includes note heads and beams. In I_{BL} image, all the beams which belong to each group symbols have been converted to a unique line. SE is rectangle structural element for morphological closing whose size is computed by eq. 15.

$$SE = \left[\frac{space\ between\ two\ line}{2} + thichness\ of\ staff\ line, \frac{space\ between\ two\ line}{2} + thichness\ of\ staff\ line\right] \tag{15}$$

The output image of morphological closing (I_{BL}), includes note heads and unique lines. We need only the unique lines to compute the inclination(angle) and slope of the beam. Therefore,

Fig. 14 Morphological closing on I_{beam} in order to convert beams to unique line

Fig. 15 Separating unique line from other objects in I_{BL}. A is the feature matrix and the bottom image is the I_{UL}



# of elements	Width	Height
1	14	8
2	78	42
3	14	10
4	14	1
6	15	14
7	14	9

feature matrix(A)
max(A(:,2))

2	78	42

the note heads should be separate from unique lines. For this reason, we utilize connected component technique to extract some simple features such as height and width of component in order to isolate them. The unique lines have the longest width in comparison to the others. Therefore, the maximum width indicates the unique lines. The output of this step is called I_{UL}. The result is shown in Fig. 15.

For obtaining the edge of the line in I_{UL}, the Laplacian filter is applied on it. The output of this step is called I_{edge}. The result of this step is shown in Fig. 16.

Finally, Hough Transform technique will be applied on the result of the previous step to detect lines according to eq. 16. Hough Transform returns three outputs; the first is RHO which is distance from the origin to the line along a vector perpendicular to the line, the second output is Theta which is angle between the x-axis and this vector. Third, H is a matrix of parameter

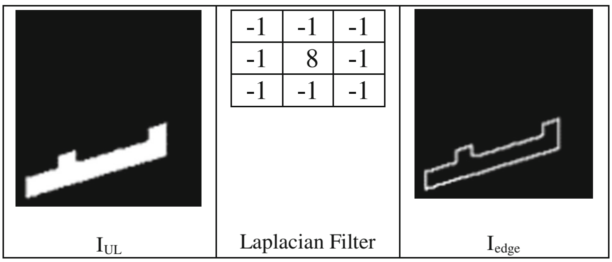


Fig. 16 Applying Laplacian filter, left image is I_{UL}. The middle image is Laplacian filter and the right image is I_{edge}

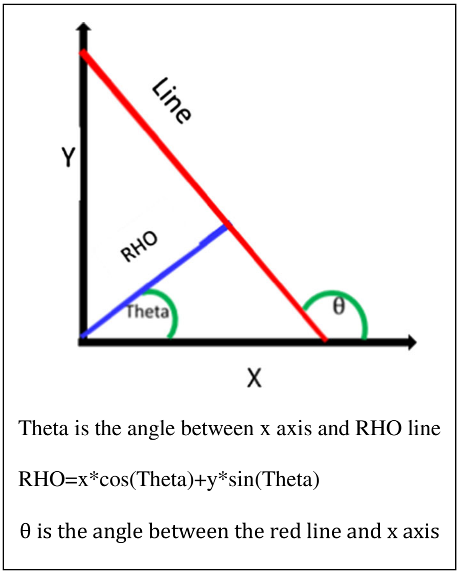


Fig. 17 The parameters of red line in parameter space

space whose rows and columns indicate the Theta and RHO values respectively. The parameters of line are shown in Fig. 17.

$$[\mathrm{H}.Theta.\mathrm{RHO}] = hough\left(\mathrm{I}_{edge}\right) \tag{16}$$

The outputs of eq. 16 are used in eqs. 17 and 18 to obtain the candidate lines in $\mathrm{I}_{\mathrm{edge}}$

$$peaks = houghpeaks(\mathrm{H}, \mathrm{n}); \tag{17}$$

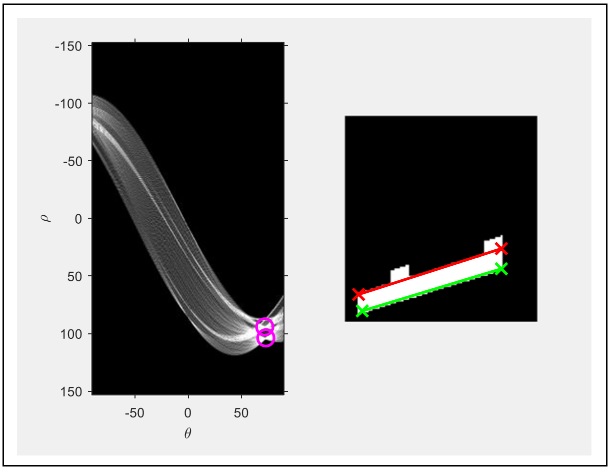
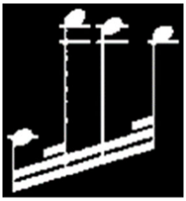
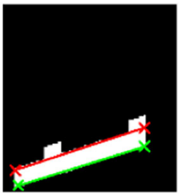


Fig. 18 The left image represents the extracted candidate lines in parameter space and the right image shows the candidate lines in I_{UL} image

a



b



c

# of Lines	Beginning point	End point	Theta	RHO	Length of line
1	[10,107]	[86,84]	73	104	79.40
2	**[8,98]**	**[86,73]**	**72**	**94.50**	**81.90**

Theta=72, θ = 90 + 72

d

Fig. 19 **a** Group symbols with skewness **b** Extract candidate lines and find the longest line with its parameters **c** Use the parameters in transformation model to remove skewness

$$lines = houghlines\left(\mathrm{I}_{edge}, Theta, \mathrm{RHO}, peaks\right); \tag{18}$$

The houghpeak function is utilized to locate the peaks in hough transform matrix in eq. 17. In this formula, H is the hough transform matrix which is obtain by eq. 16 and n indicates the maximum number of peaks. The peaks are n*2 matrices which holds the rows and columns coordinates of maximum values in H. After that the output of eq. 17 is used in eq. 18 as a parameter to find candidate lines. In eq. 18, houghlines function is used to extract candidate lines. The lines are the n*5 structure arrays which hold the features of the extracted lines. They are the length, the beginning and end points of the extracted lines, RHO, and Theta. Fig. 18. shows the candidate lines in parameter space and I_{UL} image.

Finally, the line with maximum length is chosen. After that, θ, the angle between the chosen line and x axis is obtained by eq. 19 and it is used in eq. 12 to remove skewness of the group symbols. (see Fig. 19.)

$$\theta = 90 + Theta\ of\ \ the\ longest\ line \tag{19}$$

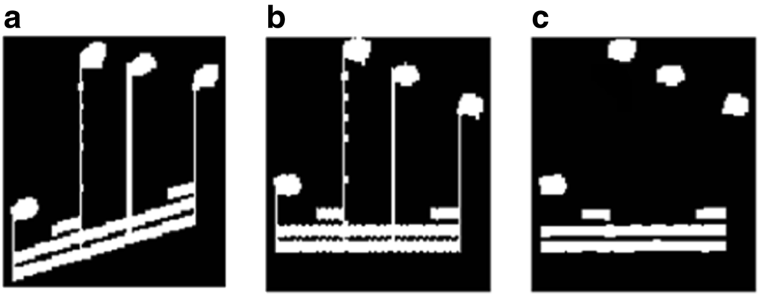


Fig. 20 **a** The group symbols with skewness **b** Removing skewness **c** Separating the note heads and beams by using opening filter

Identifying and labeling the group symbols After changing images to horizontal shape, by using morphological opening filter, the vertical line is removed and the note heads are separated from the beams. The type of structure element is linear and its size is twice the thickness of the staff lines plus one. The image output of this step is I $_{[\text{beam note}]}$, which is shown in Fig. 20c.

Now, we apply connected component algorithm on I $_{[\text{beam note}]}$ image. The output of step is called I_{CC}. For recognizing the elements of group symbols, such as note heads and beams, first

I_{CC} image

feature matrix A

# of elements	Width	Height
cc1	10	8
cc2	81	5
cc3	81	5
cc4	12	4
cc5	11	11
cc6	11	7
cc7	12	4
cc8	10	10

Image result after using thresholds

I_B

# of beam	Width	Height
cc2	81	5
cc3	81	5
cc4	12	4
cc7	12	4

I_N

# of note head	Width	Height
cc1	10	8
cc5	11	11
cc6	11	7
cc8	10	10

Fig. 21 The process of classifying note head and beam symbols

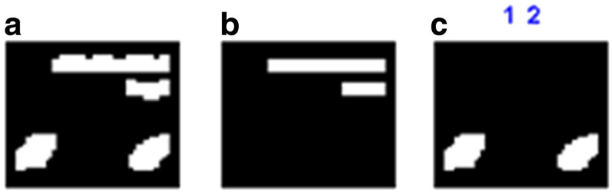


Fig. 22 **a** Removing the vertical lines **b** Beam detection **c** Labeling symbols (1 = the Eighth Note, 2 = Sixteenth Note)

the feature matrix A is extracted from I_{CC}. It includes two important features; width and height of each object in the image. After that, we will see that the note heads are the same width and height, but this is not true about the beams. The beam's width is at least twice its height and usually more. By using a simple feature like aspect ratio (width to height ratio) and comparing with a threshold, the aspect ratio can easily separate the long beam but it is not true about small beams. Therefore, after using aspect ratio to detect long beams, we use two important features to detect small beams from note heads. They are created from the thickness of staff lines and the vertical distance between two lines. We use them in eq. 20 to identify the beams and note heads symbols.

$$\textit{for each element } i \textit{ in } I_{CC} \begin{cases} i \in I_B & \left(\dfrac{width_i}{height_i}\right) > 2^{*} \textit{ thickness of staff line} \\ i \in I_B & \left(\dfrac{width_i}{(\textit{vertical distance of staff line} + 2^{*} \textit{ thickness of staff line}) - 1}\right) > 0.4 \\ i \in I_B & \left(\dfrac{height_i}{(\textit{vertical distance of staff line} + 2^{*} \textit{ thickness of staff line}) - 1}\right) > 0.4 \\ i \in I_N & \textit{Otherwise} \end{cases} \tag{20}$$

In eq. 20, I_B is an image which contains just the beam symbols, I_N shows the note head image. Fig. 21. represent the steps of this process.

Therefore, the type of note symbols can be identified by calculating the center of each note head and then counting the beams which are placed right above or below the center of the corresponding note heads.

To label the grouping symbols, a set of positive numbers is used. For example, the Eighth Note with 1, and Sixteenth Note with 2, and Thirty Second Note with 3(see Fig. 22). These numbers indicate the number of beams. A more complex example is shown in (Figs. 23, 24, 25).

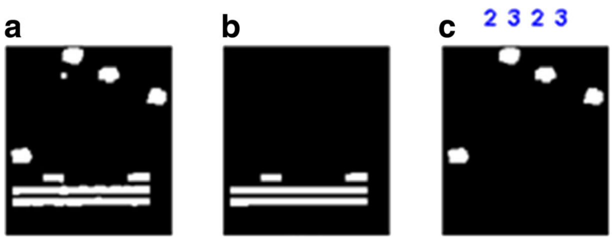


Fig. 23 **a** Removing the vertical lines **b** Beam detection **c** Labeling symbols (2 = Sixteenth Note, 3 = Thirty-second Note

Fig. 24 **a** Removing the vertical lines **b** Beam detection **c** Labeling symbols (2 = Sixteenth Note, 3 = Thirty-second Note, 4 = Sixty-fourth Note)

4 Results

In this part, the results of the proposed method on database are discussed and this method is then compared with some methods that work on this database.

4.1 Database

The databases which are used in the first phase to remove the staff lines are CVC-MUSCIMA[2] and ICDAR 2013.[3] The first is collected by Alicia Fornts et al. [13]. It includes 1000 musical sheets, which are handwritten and composed by 50 composers in the same conditions. For measuring the strength and robustness of the proposed algorithm, and to compare with other methods, the distortion model which was proposed by Dalitz et al. [10] was used. The distortion images are Ideal, Kunongo, Staff Line Thickness Variation, Staff line-y-variation, White speckles, Typeset-emulation, Thickness-ratio, Interrupted, Curvature and Rotation. After obtaining the distortion images, the number of images in database was increased to 12,000 images.

The second database is ICDAR 2013 which was introduced in [14, 47]. It includes two new kind of degradations: local noise and 3D distortion which were generated based on original CVC MUSICMA.

The database[4] which is used in the second and third phases, was introduced by Dalitz et al. [10]. It includes 32 printed sheet music images and it consists of a variety of music forms such as common music notation, lute tablature, chant and mensural notation. In addition, it covers the wide range of music fonts. The labels for musical symbols were produced manually. All the databases have only the basic ground truth for staff line removal.

4.2 Staff line removal

For removing the staff lines, the databases, CVC-MUSCIMA and ICDAR 2013 have been used, and for evaluation, the criteria, such as Precision (Positive Predictive Value), Recall (True Positive Rate or Sensitivity), Error Rate, Specificity (True Negative Rate), F-Measure and Accuracy (Classification Rate) have been computed and used as follows:

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad (21)$$

[2] http://www.cvc.uab.es/cvcmuscima/index_database.html
[3] http://dag.cvc.uab.es/muscima/
[4] http://gamera.informatik.hsnr.de/addons/musicstaves/

Fig. 25 **a** Removing the vertical lines **b** Beam detection **c** Labeling symbols (1 = the Eighth Note, 2 = Sixteenth Note, 3 = Thirty-second Note)

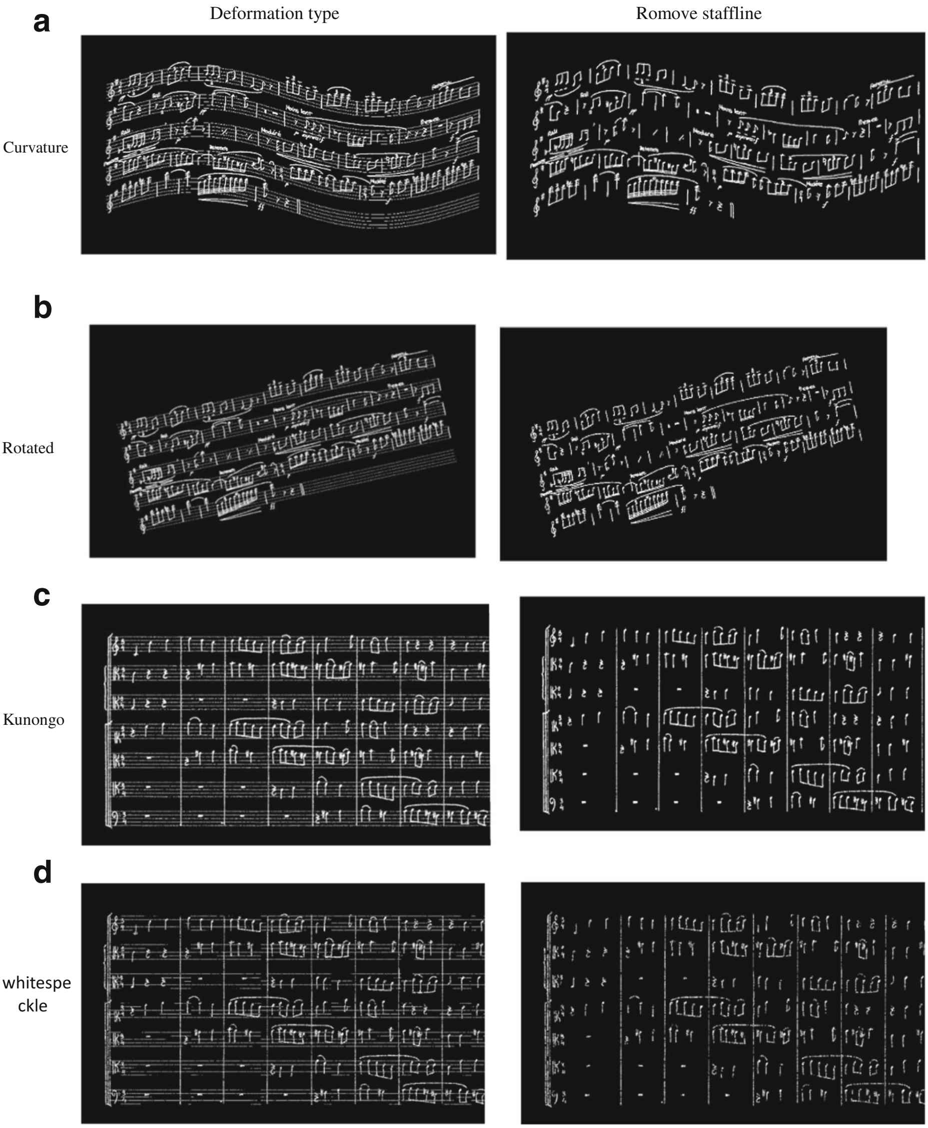


Fig. 26 Results of the proposed method to remove staff lines on four different modes of distortion (**a**) curvature, (**b**) rotation, (**c**) Kunongo, (**d**)whitespeckle. The left column indicates the original image and the right column shows the image after removing staff lines

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{22}$$

$$ERROR = 100.\frac{|missclassified\ \mathrm{SP}| + |missclassified\ \mathrm{NSP}|}{|\mathrm{all\ SP}| + |\mathrm{all\ NSP}|} \tag{23}$$

$$\mathrm{F}{-}Measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{24}$$

$$Specificity = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{25}$$

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{26}$$

In the above equations, TP is the number of pixels correctly classified as staff lines. FP is the number of pixels incorrectly considered as staff lines. FN is the number of pixels that are wrongly considered as anything other than a staff line. TN (True Negative pixels) is number of pixels correctly classified as non-staff lines. In eq. (23), misclassified SP indicates pixels that are wrongly considered as staff lines, and misclassified NSP shows pixels that are wrongly considered symbols other than staff lines. We utilized Error Rate, Precision and Recall to evaluate proposed method on CVC-MUSCIMA database. The results of the proposed method on CVC-MUSCIMA database are shown in the four states of curvature, rotation, Kunongo, and White speckles in Fig. 26.

Table 2 shows the results of the proposed method on all forms of distortion in CVC-MUSCIMA database. The configurations have been adjusted upon [15]. In all cases the error is very low. The least error rate is in Ideal, and the highest error rate is in Kunongo. Considering the Precision measure, the Ideal and Interrupted have the highest Precision, and the Kunongo has the lowest Precision. The Recall rate is the highest in the case of Ideal. Curvature, Rotation and Kunongo had the lowest Recall rate among all distortions. According to the results, our method responds very well for all kind of distortion.

To evaluate the proposed method, the average error rate is computed and compared to well know methods such as ISI01-HA which was winner in ICDAR 2011 [12] and music score competition 2012 [15]. In Table 3, you can see that the proposed method in terms of error rate is lower than all other methods. Therefore, our method outperforms the state of the art method on CVC-MUSCIMA database. In addition, it indicates that our method is robust in all kinds of distortions in sheet music images.

The result of well-known methods on ICDAR 2013 database is shown in Table 4. The Accuracy(A), Recall rate(R), Specificity(S) and F-Measure are computed in order to compare our method with other methods. In Table 4, you can see that the proposed method in Specificity measure is higher than all other methods. In addition, in Accuracy measure, our method is in the second rank among other methods. FS method [24] is the most successful method in this measure. Also, regarding the Recall rate and F-Measure, our method is the best

Table 2 Result of the proposed method in all modes of distortion on CVC-MUSCIMA database

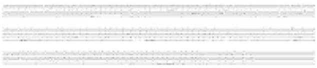
Shape	Distortion Type	Error Rate	Precision	Recall
	Ideal	1.31	97.88	97.53
	Kunongo	2.80	96.29	95.83
	Staff line Thickness Variation	1.43	97.69	97.20
	Staffline-y-variation	1.37	97.63	97.38
	White speckles	1.48	97.78	97.25
	Typeset-emulation	1.63	97.73	97.18
	Thickness-ratio	1.53	97.81	97.21
	Interrupted	1.39	97.83	97.12
	Curvature	1.71	96.55	95.74
	Rotated	2.51	96.75	95.87
	Average	1.72	97.39	96.83

Table 3 Comparison of proposed method with other methods on CVC-MUSCIMA database

Staff Removal Methods	Error Rate
ISI01-HA [12, 15]	1.89
Timofte et al. [44]	1.76
Our method	**1.72**

Table 4 Comparison of proposed method with well-known methods on ICDIR 2013 dataset

Method	A	R	S	F
LTC [10] (2008)	87.58	64.66	99.53	80.40
Skeleton [10] (2008)	94.50	86.97	99.03	92.24
LRDE [18] (2014)	97.03	94.02	98.84	95.97
FS [24] (2014)	**97.34**	**95.87**	98.22	95.90
Manual [38] (2014)	96.89	94.37	98.41	95.51
NILC [25] (2016)	97.10	94.58	98.61	**96.07**
KA [26] (2016)	96.23	92.85	98.26	94.88
Our method	97.21	95.18	**99.71**	**96.06**

method after the FS and NILC method [25] respectively. Generally, our method is the best after FS method in Accuracy and Recall rate metric and it is only slightly less than FS method in Accuracy, Recall and F-Measure. Although, our method is the first rank in term of specificity. Finally, we could claim our method is comparable with the best existing methods.

4.3 Musical symbols recognition

First of all, we classified the musical symbols into two groups; group symbols, and single symbols. The single symbols can appear in musical sheets separately. A combination of single symbols creates group symbols. Therefore, they have to be identified and separated from single symbols in order to be analyzed and recognized their components. Some of the musical symbols are shown in Table 5. In a test conducted to identify musical symbols, the database provided by Dulitz et al. [10] is used. This database contains 32 idealized sheets from musical images. Database prepared by [34] is used to obtain the best pattern. Criteria selected for

Table 5 Group and single musical symbols

Shape	Single symbols	Can be combined with each other (group symbols)
♯	Sharp	-
𝄠	Treble Clef	-
𝅗𝅥	Half Note	-
♪	-	Eighth Note
𝅘𝅥𝅯	-	Sixteenth Note
𝅘𝅥𝅰	-	Thirty second Note
𝅘𝅥𝅱	-	Sixty-fourth Note and other with more flags

Table 6 Result of classification of musical symbols

Shape	Symbols	Total number	True	False	Error Rate
♯	Sharp	181	178	3	0.017
♭	Flat	141	137	4	0.028
♮	Natural	65	61	4	0.062
𝄞	Treble Clef	76	71	5	0.066
𝄢	Bass Clef	32	28	4	0.125
𝄡	Alto Clef	16	10	6	0.375
𝄼	Half Note rest	72	66	6	0.083
𝄽	Quarter Note rest	49	42	7	0.143
𝄾𝄿	Eighth and Sixteenth Note rest	101	93	8	0.079
·	Dot	278	268	10	0.036
𝅝	whole Note	78	75	3	0.039
𝅗𝅥	Half Note	381	376	5	0.013
𝅘𝅥	Quarter Note	638	629	9	0.014
𝅘𝅥𝅮	Eighth Note	600	580	10	0.017
𝅘𝅥𝅯	Sixteenth Note	588	570	18	0.030
𝅘𝅥𝅰	Thirty-second Note	220	199	21	0.096
𝅘𝅥𝅱	Sixty-fourth Note and others with more flags	38	28	10	0.263
17	Mean Error				0.0874

algorithm evaluation were based on the number of correct data for each class. Musical symbol labels have been performed manually.

The results are shown in Table 6. This shows the amount of classified error for each musical symbol. After removing the staff lines, 17 classes were considered for the remaining musical symbols. Four of these 17 relate to the analyzed components of group symbols, which consist of Eighth Note, Sixteenth Note, Thirty-second Note, and Sixty-fourth Note and the others with more flags. Because the identification of group symbols is important, our focus in this phase is on recognition of group symbols. To see the efficiency of the proposed method, the mean absolute errors of each class (0.08%) was obtained, which is quite acceptable. The strength of the proposed method is on the detection and analysis of group symbols, because it is quite robust to skewness of group symbols and has very high accuracy in identifying symbols.

5 Conclusion

In this paper, we proposed a method to analyze sheet music images. In the first phase of the method, a robust algorithm for removing staff lines in all cases which were defined in CVC-MUSCIMA and ICDIR 2013 databases, was proposed. Indeed, the focus was on handwritten musical sheets. In comparison to other methods on CVC-MUSCIMA, the method had the lowest error rate, 1.72. Also in ICDAR 2013 database, our specificity was 99.71% which performs the best among other methods. Moreover, the accuracy, recall and f-measure of our method is placed in second rank and it is just slightly less than the first rank method on this database. In the second phase, a simple method was proposed in order to detect musical symbols and divided them into two categories; single and group symbols. The final phase was the recognition of musical symbols. For this phase, two methods were proposed: a pattern matching method for identifying single symbols and a hierarchical algorithm to recognize group symbols. The major focus of the proposed method in this phase was on the recognition of group symbols. In the first phase, the proposed method works well on removing the staff lines of handwritten pages, however in the second and third phases, with the aim of recognition of the musical symbols, it does not work as well as the first phase, but the overall results are satisfactory. In the future, we can work on recognition of musical symbols on handwritten sheets, and also focus on online methods which recognize musical symbols. Then it is possible to use them in devices such as a phones, tablets, or glasses such as Google glasses. This method is important because it will ease teaching the music, especially to those who are visually impaired or blind. Moreover, it is also possible to compose music by using a series of cameras which track the finger movements of a piano player and detect all musical symbols based on his/her fingers and write it on a musical sheet.

References

1. Bainbridge D, Bell T (1996) An extensible optical music recognition system. Aust Comput Sci Commun 18:308–317
2. Bellini P and Nesi P (2001) WEDELMUSIC format: An XML music notation format for emerging applications, In Web Delivering of Music, 2001. Proceedings. First International Conference on, pp. 79–86
3. Blostein D and Baird HS (1992) A critical survey of music image analysis, In Structured Document Image Analysis, Springer, pp. 405–434
4. Calvo-Zaragoza J and Oncina J (2014) Recognition of Pen-Based Music Notation: the HOMUS dataset, In Pattern Recognition (ICPR), 2014 22nd International Conference on, pp. 3038–3043
5. Calvo-Zaragoza J et al (2015) Avoiding staff removal stage in optical music recognition: application to scores written in white mensural notation. Pattern Anal Applic 18(4):933–943
6. Calvo-Zaragoza J, Mico L, and Oncina J (2016) Music staff removal with supervised pixel classification International Journal on Document Analysis and Recognition (IJDAR), pages 1–9
7. Calvo-Zaragoza J, Vigliensoni G, and Fujinaga I (2016) STAFF-LINE DETECTION ON GREYSCALE IMAGES WITH PIXEL CLASSIFICATION
8. Carter NP (1989) Automatic recognition of printed music in the context of electronic publishing." University of Surrey

9. Coüasnon B and Camillerapp J, (1994) Using grammars to segment and recognize music scores, In International Association for Pattern Recognition Workshop on Document Analysis Systems, pp. 15–27
10. Dalitz C, Droettboom M, Pranzas B, Fujinaga I (2008) A comparative study of staff removal algorithms. IEEE Trans Pattern Anal Mach Intell 30(5):753–766
11. Dutta A, Pal U, Fornes A, and Llados J (2010) An efficient staff removal approach from printed musical documents, In Pattern Recognition (ICPR), 2010 20th International Conference on, pp. 1965–1968
12. Fornes A, Dutta A, Gordo A, and Llados J (2011) The ICDAR 2011 music scores competition: Staff removal and writer identification. In Document Analysis and Recognition (ICDAR), 2011 International Conference on, pp. 1511–1515. IEEE
13. Fornés A, Dutta A, Gordo A, Lladós J (2012) CVC-MUSCIMA: a ground truth of handwritten music score images for writer identification and staff removal. Int J Doc Anal Recognit 15(3):243–251
14. Fornés, A, Kieu VC, Visani M, Journet N, and Dutta A (2013) The ICDAR/GREC 2013 music scores competition: Staff removal. In International Workshop on Graphics Recognition, pp. 207–220. Springer Berlin Heidelberg
15. Fornés, A, Dutta A, Gordo A, and Lladós J (2013) The 2012 music scores competitions: staff removal and writer identification. In Graphics Recognition. New Trends and Challenges, pp. 173–186. Springer Berlin Heidelberg
16. Fujinaga I (2004) Staff detection and removal, Vis. Percept. Music Not. on-line off-line Recognit., pp. 1–39
17. Genfang C, Liyin Z, Wenjun Z, and Qiuqiu W (2010) Detecting the staff-lines of musical score with hough transform and mathematical morphology, In Multimedia Technology (ICMT), 2010 International Conference on, pp. 1–4
18. Géraud, T (2014) A morphological method for music score staff removal." 2014 I.E. International Conference on Image Processing (ICIP). IEEE
19. Gonzalez RC, Woods RE, and Eddins SL (2004) Digital image processing using MATLAB. Pearson Education India
20. Hirata NST (2009) Multilevel training of binary morphological operators. IEEE Trans Pattern Anal Mach Intell 31(4):707–720
21. Luangnapa N, Silpavarangkura T, Nukoolkit C, and Mongkolnam P (2012) Optical Music Recognition on Android Platform, In *Advances in Information Technology*, Springer, pp. 106–115
22. Mahoney JV (1982) Automatic analysis of music score images. Massachusetts Institute of Technology
23. Miyao H, Okamoto M (2004) Stave Extraction for Printed Music Scores Using DP Matching. JACIII 8(2): 208–215
24. Montagner, IS, Hirata R, and Hirata NST (2014) Learning to remove staff lines from music score images. In Image Processing (ICIP), 2014 I.E. International Conference on, pp. 2614–2618. IEEE
25. Montagner IS, Hirata NST, Hirata R, and Canu S (2016) NILC: A two level learning algorithm with operator selection. In Image Processing (ICIP), 2016 I.E. International Conference on, pp. 1873–1877. IEEE
26. Montagner IS, Canu S, Hirata NST, and Hirata R Jr, (2016) Kernel Approximations for W-operator learning," in to appear in Proceedings of SIBGRAPI 2016 (XXIX Conference on Patterns, Graphics and Images)
27. Ng K (2004) Optical music analysis for printed music score and handwritten music manuscript, Vis. Percept. Music Not. On-Line Off-Line Recognit., pp. 108–127
28. Nhat VQ and Lee G (2014) Adaptive line fitting for staff detection in handwritten music score images," In Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication, p. 99
29. Prerau DS (1970) Computer pattern recognition of standard engraved music notation," Massachusetts Institute of Technology. Department of Electrical Engineering. 1970. Ph. D.
30. Pruslin DH (1966) Automatic recognition of sheet music. Massachusetts Institute of Technology
31. Pugin L, (2006) Optical Music Recognitoin of Early Typographic Prints using Hidden Markov Models. In ISMIR, pp. 53–56
32. Ramirez C, Ohya J (2014) Automatic recognition of square notation symbols in western plainchant manuscripts. J New Music Res 43(4):390–399
33. Randriamahefa R, Cocquerez JP, Fluhr C, Pepin F, and Philipp S (1993) Printed music recognition, In Document Analysis and Recognition, 1993., Proceedings of the Second International Conference on, pp. 898–901
34. Rebelo A, Capela G, Cardoso JS (2010) Optical recognition of music symbols. Int J Doc Anal Recognit 13(1):19–31
35. Rebelo A, Paszkiewicz F, Guedes C, Marcal ARS, and Cardoso J (2011) A method for music symbols extraction based on musical rules, In Proceedings of Bridges 2011: Mathematics, Music, Art, Architecture, Culture, pp. 81–88

36. Rossant F, Bloch I (2007) Robust and adaptive OMR system including fuzzy modeling, fusion of musical rules, and possible error detection. EURASIP J Appl Signal Process 2007(1):160
37. dos Santos Cardoso J, Capela A, Rebelo A, Guedes C, Pinto da Costa J (2009) Staff detection with stable paths. IEEE Trans Pattern Anal Mach Intell 31(6):1134–1139
38. dos Santos Montagner I, Hirata R, and Hirata NST (2014) A machine learning based method for staff removal." Pattern Recognition (ICPR), 2014 22nd International Conference on. IEEE
39. Soontornwutikul T, Thananart N, Wantanareeyachart A, Nukoolkit C, and Arpnikanondt C (2013) Optical Music Recognition on Windows Phone 7, In The 9th International Conference on Computing and InformationTechnology (IC2IT2013), pp. 239–248
40. Sotoodeh M and Tajeripour F (2012) Staff detection and removal using derivation and connected component analysis, In Artificial Intelligence and Signal Processing (AISP), 2012 16th CSI International Symposium on, pp. 54–57
41. Su B, Lu S, Pal U, and Tan CL, (2012) An effective staff detection and removal technique for musical documents, In Document Analysis Systems (DAS), 2012 10th IAPR International Workshop on, pp. 160–164
42. Szwoch M (2005) A robust detector for distorted music staves, In computer analysis of images and patterns, pp. 701–708
43. Tajeripour F, Sotoodeh M (2012) A novel staff removal method for printed music image. IEICE Electron Express 9(7):609–615
44. Timofte, R, and Van Gool L (2012) Automatic stave discovery for musical facsimiles." In Asian Conference on Computer Vision, pp. 510–523. Springer Berlin Heidelberg
45. Toyama F, Shoji K, and Miyamichi J (2006) Symbol recognition of printed piano scores with touching symbols, In Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, vol. 2, pp. 480–483
46. Tsandilas T, (2012) Interpreting strokes on paper with a mobile assistant, In Proceedings of the 25th annual ACM symposium on User interface software and technology, pp. 299–308
47. Visaniy, M, Kieu VC, Fornés A, and Journet N (2013) Icdar 2013 music scores competition: Staff removal." In Document Analysis and Recognition (ICDAR), 2013 12th International Conference on, pp. 1407–1411. IEEE

Mahmood Sotoodeh received his B.Sc. degree in Computer Engineering from Shiraz University, Shiraz, Iran, in 2008, and his M.Sc. degree in Artificial Intelligence from Shiraz University, Iran in 2012. He attained the second rank in Blind and low-sighted Empowerment Conference and Scientific Research. He worked for six months as a research scholar in Utah State University, Utah, USA. Currently, he is PhD candidate in Shiraz University. His research interests include image processing, computer vision and pattern recognition,

Farshad Tajeripour received the B.S. and M.S. degrees in Electronic engineering from Shiraz University, Shiraz, Iran, in 1994 and 1997. He received Ph.D. degree in Electronic engineering from Tarbiat Modarres University, Tehran, Iran, in 2009. Currently he is an assistant professor in Shiraz University. His research interests include texture classification, pattern recognition, computer vision and video processing.

Sadegh Teimori received his B.Sc. degree in software Engineering from Islamic Azad University of Shiraz, Shiraz, Iran, in 2010, and his M.Sc. degree in Artificial Intelligence from Shiraz University, Iran in 2014. His research interests include image processing, computer vision and pattern recognition.

Kirk Jorgensen received his Bachelor of Science and Art degree from Utah State University, USA in 1983. He received his Associate degree of Psychology from Ricks College, Rexburg Idaho, USA in 1976. He is currently an instructor at the English Language Center of Logan Utah and an instructor of piano music at the Main Street Music Academy also in Logan Utah.