

# Optical Music Score Recognition System for Smart Mobile Devices

SeJin Han, GueeSang Lee\*

Department of Electronics and Computer Engineering  
Chonnam National University, Gwangju, 500-757, Korea

## ABSTRACT

*In this paper, we propose a smart system that can optically recognize a music score within a document and can play the music after recognition. Many historic handwritten documents have now been digitalized. Converting images of a music score within documents into digital files is particularly difficult and requires considerable resources because a music score consists of a 2D structure with both staff lines and symbols. The proposed system takes an input image using a mobile device equipped with a camera module, and the image is optimized via preprocessing. Binarization, music sheet correction, staff line recognition, vertical line detection, note recognition, and symbol recognition processing are then applied, and a music file is generated in an XML format. The Music XML file is recorded as digital information, and based on that file, we can modify the result, logically correct errors, and finally generate a MIDI file. Our system reduces misrecognition, and a wider range of music score can be recognized because we have implemented distortion correction and vertical line detection. We show that the proposed method is practical, and that it has potential for wide application through an experiment with a variety of music scores.*

**Key words:** Music Recognition, Music OCR, Optical Music Score Recognition.

## 1. INTRODUCTION

Many paper documents have now been converted to digital files. The conversion mainly works through optical recognition and many studies have been conducted to improve its efficiency. Digitalizing a music score document is a considerable challenge because music scores consist of structures that are complex and difficult to recognize. Symbols connected to five staff lines are very difficult to digitalize and are time consuming. Therefore, many systems have been designed and developed. In particular, a system is needed that can be utilized in a mobile device.

Mobile device hardware and software performance have continued to be developed. Nowadays, people can undertake similar tasks on a mobile phone to those on a PC. Thus, users desire more useful applications and a music score recognition system is needed.

In this paper, we suggest a music score recognition system for a mobile device such as a smart phone. First, because smart phones have a built-in camera module, a music score document can be captured by the camera and the image quality can be improved and blank areas removed in the preprocess stage. After preprocessing, we apply an advanced binarization method to the image. Generally, we minimize the noise in the music score documents for the experiment

and then take a picture on the flat ground. However, it is difficult to obtain a noiseless image in real life. Most music scores are found within bound textbooks or hymnals. Thus, our optical recognition system needs distortion correction before staff line detection. On the other hand, when we attempt binarization or the removal of the staff line in a low quality image, some parts of the symbols become damaged. Our OMR system therefore removes the text or noise such as lyrics and conducts vertical line detection. Vertical line detection is an important algorithm in this system. Most key symbols in music scores have vertical lines that have their own position and length. We can recognize most notes and symbols using vertical line detection. Finally, an XML file and Midi file are generated to display and play the song. Fig. 1 shows the entire configuration for our system.

We present the detail contents in each section. Section 2 discusses the input phase, section 3 discusses the recognition phase, and the output phase is explained in section 4. Section 5 shows the experiment results. Lastly, we discuss some important findings and remaining challenges in the conclusion.

---

\* Corresponding author, Email: [gslee@jnu.ac.kr](mailto:gslee@jnu.ac.kr)

Manuscript received Sep. 01, 2014; revised Dec. 05, 2014;  
accepted Dec. 15, 2014

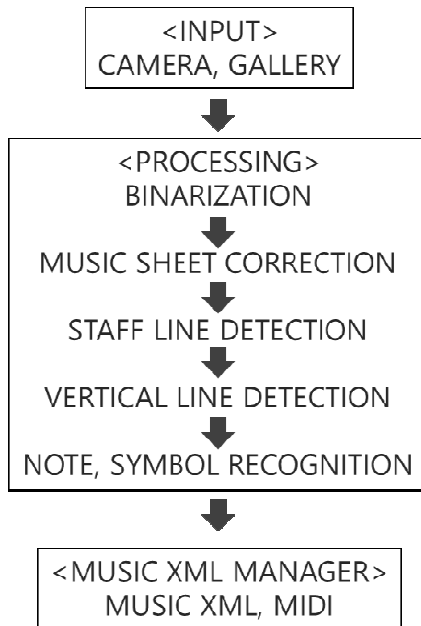


Fig. 1. System Configuration

## 2. INPUT OF SYSTEM

Our system has 3 phases: input, processing, and manager, as shown in fig. 1. In the input phase, a 'Camera' and a 'Gallery' are used. The camera is the most important component of the system because the main feature of the mobile device is portability, since it can be used anywhere and anytime. We thus need to pay particular attention to the camera. Furthermore, the image can get by other camera module or different time. Therefore, our system supports the gallery function.

### A. Camera

Our OMR system first captures a picture with the camera module. When the application is loaded, the camera menu is first selected to obtain an input. The system shows a preview display with a blue guideline for the optimal image. The guideline helps users to capture a good image because it can check the perspective and angle of rotation. Therefore, when the user touches the shutter button, the system uses this picture as the input image. In general, the mobile camera is directly affected by the surrounding light and environment. Attention needs to be given to the surroundings unexpected changes in image quality. Our system uses 'text mode', and can handle a high contrast between a black object symbol and white background. Thus, our system can efficiently apply binarization processing to the image.

### B. Gallery

Another advantage of the camera in our system is that an image can be loaded from the gallery. The system finds images from both internal and external storage and provides thumbnails and file information briefly for user convenience. The input type is in jpeg format and the processing sequence

is the same as the camera mode. Fig. 2 shows the gallery interface.

## 3. RECOGNITION MODULE

After the input phase, the system processes recognition of the music score. Color information is removed because processing based on color information involves too much exception. In order to use form and location information, the system needs a binarized image. Despite the camera module's guideline, almost all images

have some distortions. We need to correct the staff lines in the distorted image with a distortion correction module.

The main function of recognition is vertical line detection. A music sheet has many notes and symbols, most of which include more than one vertical line. We recognize the music score using the result of the vertical line and staff line detection.



Fig. 2. Gallery Interface

### A. Binarization

Instead of color information, our system uses form and location information for recognition. The recognition section receives a bitmap as a parameter from the input phase. As described previously, a mobile camera is strongly affected by external components, while factors such as the distance between the lens and the document, the gradient of the lens, and illumination can change unexpectedly. The binarization method of our system should deal effectively with these issues. Our system uses the advanced binarization method suggested by J.M. Yoo et al. [1]. This method calculates the global threshold and local threshold value for each block. It is also robust to the light variation condition at the time the image is taken. Many well-known binarization methods have been used in the study of image processing, but the advanced binarization method is specialized to a music score image.

**B. Distortion correction**

Before the recognizing stage, the input image has some problems. When we take a picture, we need to consider the perspective of the lens, the rotation of the music sheet, and the curve of the page when the actual book is open. Our system employs the biquadratic transformation model [6] to solve these issues. Fig. 3 shows an example of a distorted image and the correct image.

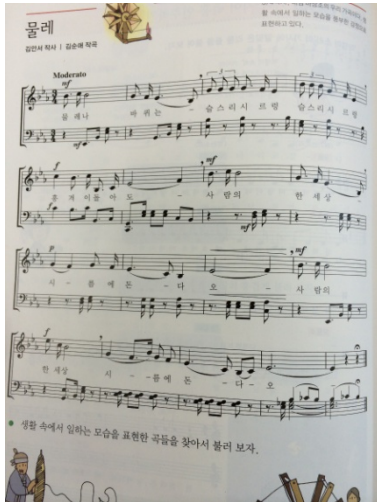


Fig. 3-(a)



Fig. 3-(b)

Fig. 3. (a) Distorted image, (b) Corrected image

**C. Staff line detection**

A staff line consists of five long thin horizontal lines with an interval between each line. Various symbols and music notes are located on these lines in various positions. Our system uses the adaptive line fitting method [2] to detect the staff line. It can overcome the curved or disconnected line problem in the worst quality image. We can obtain the number of staff line groups, the position of each line, the distance between the inside staff lines, and the thickness of each line. This is an important tool for musical symbol recognition. Fig. 4 shows an example of a disconnected staff line image and the result of our method.

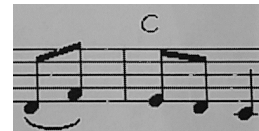


Fig. 4-(a)

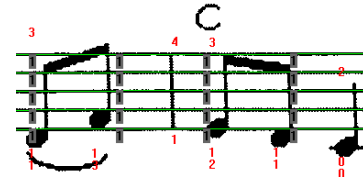


Fig. 4-(b)

Fig. 4. Staff line detection (a) Disconnected line, (b) Result image

**D. Vertical line detection**

In the previous OMR system, each symbol is separated from the image and the features are extracted. However, if a low quality image is derived from the binarization module, the symbol is broken. It is also difficult to discriminate the target symbol from the other symbols.

Also, in the previous OMR system, the staff line was removed for recognition, causing a similar problem of the loss of the original shape. In this paper, we use vertical line detection to resolve this issue.

Most notes and symbols include more than one vertical line. The extracted information from a vertical line shows a different characteristic for each symbol. Our system can verify the vertical line within a range around the area of the staff line without removing the staff line. The longest connected line in each vertical projection position is then determined. Therefore, we can remove the chord or lyric area and find the stem, bar line or a part of the symbol. The position of the vertical line, length, and thickness are results of vertical line detection. We can then find more vertical lines, dots or heads around the detected line.



Fig. 5. Vertical lines in music sheet

**E. Note & Symbol recognition**

Music score recognition is based on the result of staff line and vertical line detection. For each stem, we apply template matching for notes, except for whole notes. Fig. 6 shows an example of a symbol that is recognized by a vertical line. When removing the staff line, the shapes of the sharps and natural symbols are lost. Sometimes, a repeated symbol is recognized as noise. A sharp has two sequential lines and a specific position on the staff line. These two lines have the same length but the second line is higher than the first line. The repeat symbol also has two sequential lines of the same length. However, the thickness differs. In addition, the repeat has two aligned dots around the symbol. The bar line, natural, and flat symbols are recognized in a similar way.

If a component has no vertical line, we analyze its own feature, constraint, size of component or width. It can be a whole note, a dot, a rest mark, and so on from the result of recognition. On the other hand, our system recognizes the chord (e.g. B, F, G7) by analyzing its distance to the top of the staff line

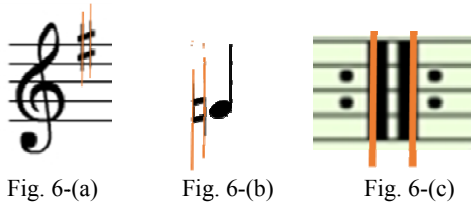


Fig. 6. Example of Symbol included on a vertical line  
 (a) Sharp mark, (b) Natural mark, and (c) Repeat mark

4. OUTPUT OF SYSTEM

Output is the final phase of our system. XML files are generated from the recognition phase. The system needs to display the files to check and modify the data. The system provides a variety of functions such as XML decoding and screen capture. This role can be performed by the Music XML manager in the output phase. The manager is the name of the new interface for our OMR system. It handles specific issues after the recognition phase.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE score-partwise SYSTEM "http://www.musicxml.org/dtds/partwise"
- <score-partwise>
- <defaults>
- <page-layout>
- <page-margins type="both">
- <left-margin>50</left-margin>
- <right-margin>50</right-margin>
- <top-margin>10</top-margin>
- <bottom-margin>50</bottom-margin>
- </page-margins>
- <page-width>1370</page-width>
- <page-height>675</page-height>
- </page-layout>
- <staff-layout>
- <staff_distance>80</staff_distance>
- </staff-layout>
- </defaults>
- <part-list>
- <part-group type="start" number="1">
- <group-symbol>bracket</group-symbol>
- <group-barline>yes</group-barline>
- </part-group>
- <score-part id="P1">
- <part-name>Piano</part-name>
- <score-instrument id="P1-I14">
- <instrument-name>Instrument 1</instrument-name>
- </score-instrument>
- <midi-instrument id="P1-I1">
- <midi-channel>1</midi-channel>
- <midi-program>1</midi-program>
- </midi-instrument>
- </score-part>
```

Fig. 7. XML Structure

A. Music XML

As described in section 3, our system determines the main and subcategory, and recognizes the symbol and location of the music note. It then makes an output as Music XML. Music XML is organized with a strict structure by an XML standard. The basic information such as the title, composer, and size of the image is in the upper class. The left and right margins, code, location and pitch, and tempo are specified in the XML file created in the recognition phase. Our system decodes the XML file and displays a digital

music score. In this interface, the user can add or modify a symbol and save again. Also, the digital music score can be captured by a screen image file. Fig. 7 shows the structure of the Music XML file.

B. Midi

The Midi file is created based on Music XML. The user can play the song in the digital interface. If the user modifies the digital music score, it is applied to the midi file and xml file. Basically, the midi file can be played by any other commercial music player.

5. EXPERIMENT RESULTS

For the experiment, ‘Samsung Galaxy Note 3’ is used and implemented in our system. In order to check the behavior of the system, a picture is directly taken by this device. Fig. 8-(a) shows the original music score document used for this experiment.

Fig. 8 shows the main screen, captured music score image, and the result of each recognition. The gallery menu shown in fig. 8-(b), as described previously, works in the same way as the camera menu. From the camera menu, the music score is captured directly and recognition processing is conducted. Our system generates a digital music score interface by analyzing the Music XML file, as shown in fig. 8-(c). The digital music score interface is generated using Music XML analyzation, which can be used to play the song, modify the symbol, and save to the midi or xml file.

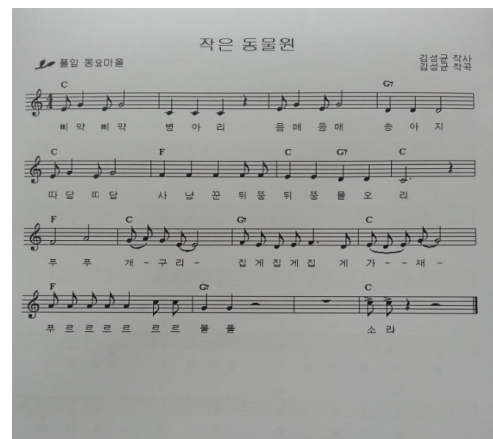


Fig. 8-(a)



Fig. 8-(b)



Fig. 8-(c)

Fig. 8. System Test. (a) Original Image, (b) Main Screen, (c) Music XML manager

## 6. CONCLUSION

In this paper, we suggest an optical music score recognition system for smart mobile devices. The system can obtain a high resolution image from the camera module, and can then conduct binarization, music sheet correction, staff line detection, and vertical line detection.

The system can then recognize the symbols and music notes using the information of the staff line and vertical line. Information such as pitch, tempo, etc. is used for generating Music XML displayed by Music XML Manager. Manager shows the digital music score and plays the midi file. Also, it can modify symbols or music notes and apply to new midi and xml files. Our OMR system can recognize the distorted image from the textbook and bent ground. It can also reduce the loss of information by removing the staff line and using binarization.

Through this system, we can use the optical music score recognition system not only on a PC, but also in a mobile device. Also, we can check its potential for development. While our system shows improved accuracy of recognition compared to the previous system and is more practical for real images, some special symbols, in exceptional cases, remain difficult to recognize from the captured images. This is the challenge for future study.

## ACKNOWLEDGEMENT

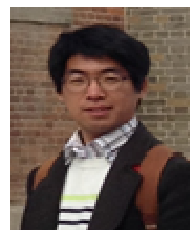
This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2014-024950) and by Samsung Electronics Co., LTD.

## REFERENCES

- [1] J. M. Yoo, N. D. Toan, D. J. Choi, H. R. Park, and G. S. Lee, "Advanced Binarization Method for Music Score

Recognition using Local Thresholds," CIT Workshops, 2008, pp. 417-420.

- [2] V. Q. Nhat and G. S. Lee, "Adaptive Line Fitting for Staff Detection in Handwritten Music Score Images," Proc. 8th ICUIMC, article no. 99, 2014.
- [3] A. Rebelo, I. Fujinaga, F. Paszkiewicz, A. R. S. Marcal, C. Guedes, and J. S. Cardoso, "Optical music recognition: state-of-the-art and open issues," International Journal of Multimedia Information Retrieval, vol. 1, issue. 3, 2012, pp. 173-190.
- [4] D. Bainbridge and T. Bell, "The Challenge of Optical Music Recognition," Computers and the Humanities, vol. 35, issue. 2, 2001, pp. 95-121.
- [5] K. H. Park, S. R. Oh, H. J. Son, J. M. Yoo, S. H. Kim, and G. S. Lee, "Decision-Tree Algorithm for Recognition of Music Score Images Obtained by Mobile Phone Camera," The Journal of the Korea Contents Association, vol. 8, no. 6, 2008, pp. 16-25.
- [6] Y. Y. Tang and C. Y. Suen, "Image transformation approach to nonlinear shape restoration," IEEE Trans. Systems, Man and Cybernetics, vol. 23, issue. 1, 1993, pp. 155-172.
- [7] T. Beran and T. Macek, "Recognition of Printed Music Score," MLDM'99, LNAI 1715, 1999, pp. 174-179.
- [8] A. Capela, J. S. Cardoso, A. Rebelo, and C. Guedes, "Integrated Recognition System for Music Scores," Proc. ICMC'2008, 2008.
- [9] A. Rebelo, G. Capela, and J. S. Cardoso, "Optical recognition of music symbols: A comparative study," IJDAR, vol. 13, issue. 1, 2010, pp. 19-31.
- [10] Q. A. Arshad, W. Z. Khan, and Z. Ihsan, "Overview of Algorithms and Techniques for Optical Music Recognition," CIIT Workshops on 4th CWRC, 2006.
- [11] K. B. Kim, W. J. Lee, and Y. W. Woo, "Automatic Recognition and Performance of Printed Musical Sheets Using Fuzzy ART," The Journal of the Korea Institute of Electronic Communication Sciences, vol. 6, no. 1, 2011, pp. 84-89.
- [12] G. S. Choudhury, T. Dilauro, M. Droettboom, I. Fujinaga, and K. Macmillan, "Strike Up the Score," D-Lib Magazine, vol. 7, no. 2, 2001.



**SeJin Han**

He received a B.S. degree in Computer Engineering from Chonnam National University, Korea in 2013. He is currently an M.S. student in the Electronics and Computer Science department of Chonnam National University, Korea. His research interests include multimedia image processing, and pattern recognition.

**GueeSang Lee**

He received a B.S. degree in Electrical Engineering and an M.S. degree in Computer Engineering from Seoul National University, Korea in 1980 and 1982, respectively. He received a Ph.D. degree in Computer Science from Pennsylvania State University in 1991.

He is currently a professor of the Department of Electronics and Computer Engineering in Chonnam National University, Korea. His research interests are mainly in the field of image processing, computer vision and video technology.