

Reading and playing musical notes with image processing techniques with mobile application

Sezgin Ersoy¹, Fatih Özdöşemeci²

¹Advanced Research in Mechatronics and Artificial Intelligence, Marmara University, Aydınevler Mah. Idealtepe Yolu No. 15 34854 Maltepe, Istanbul, Turkey

²Department of Mechatronics Engineering, Faculty of Technology, Marmara University, Aydınevler Mah. Idealtepe Yolu No. 15 34854 Maltepe, Istanbul, Turkey

¹Corresponding author

E-mail: ¹ersoy@marmara.edu.tr, ²fozyum@hotmail.com

Received 11 April 2022; received in revised form 26 April 2022; accepted 7 May 2022
DOI <https://doi.org/10.21595/vp.2022.22589>



Copyright © 2022 Sezgin Ersoy, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract. The aim of this study is to process the notes on the music sheet, written by hand on the musical note line or edited in the computer environment, with the image processing technique and to play the music over the phone application. This application will be developed with Java in the Android Studio environment and image processing will be done using the OpenCV Library. Detection of a note will be memorized by calculating the value and duration of the note. After these sounds, the frequency values of the notes will be available on the phone.

Keywords: mobil application, image processing, musical notes.

1. Introduction

Art and music have served as a visual and audio media for symbolic mass communication for thousands of years. The products of the interaction of the art and cultural environment, music and the human brain are located at the intersection of neurobiology, neurophysiology and culture [1]. Mankind first spoke by using the sounds of limited frequencies (20 Hz-20 kHz) that they could perceive, and as a result of this extraordinary communication, they turned the sound into art and revealed the phenomenon of music [2]. Sounds and images have started to increase gradually on humans, and researches and theories related to this have begun to be discussed.

Today, Art has become an industry and has gained a new dimension with technology. This field, known as digital art; It is based on computer-based digital coding or electronic storage and processing of information in different formats (text, number, image, sound) in a common binary code. The ways in which art making includes computer-based digital coding is extremely diverse. A digital photograph may be the product of a manipulated sample of visual information captured with a digital camera from a “live” scene or captured with a scanner from a conventional celluloid photograph. Music can be recorded and then digitally manipulated or created digitally with specialized computer software. “Digital art” defines the technological arts and offers many possible interpretations of fluid boundary terminology [3, 4].

Notation is a system that allows music to be recorded by transferring it to paper. In the process before the notation system was created, the works that were transferred from ear to ear were subject to change because they were not written, and disappeared after a while. For this reason, it was not possible for music to develop and spread. The discovery of sheet music made it simpler to practice music. In this way, composers were able to compose for many instruments. In addition, the notation system paved the way by universalizing music. People have come to a position where they can play pieces just by looking at the notes.

Analyzing the sounds emanating from musical instruments, determining the frequencies of them and finding their note values are of great importance for people who are far from music. With the realization of sound analysis, musical notes will be created and compositions will be performed in the computer environment. In addition to these, the analysis of the sounds in the external environment by transferring them to the computer environment and knowing their

frequencies also enable their use in voice control systems [5]. With the development of these control systems, the development of robots that can autonomously read music and sing with vocal voices [6], development of learning interfaces [7, 8], musical instrument tutorial software [9] have been developed.

In this study, the development of a mobile device application that performs the reading of musical notes with image processing technique and playing music will be presented.

This study not only performs the note reading but also provides the vocalization of the note with the image processing technique. In addition, it is presented as a contribution to the application in an application so that these processes can be applied on mobile devices.

2. Method

Music is the regular sounds consisting of the combination of sounds of a certain frequency and silence for certain periods of time. In order for these sounds to be recorded, there is a notation system that represents the frequency values and durations. The 5 lines parallel to each other that make up a line on the music sheet are called portraits or staves. The frequencies of the notes depend on the keys on the portrait, the pitch marks, and the position of the note on the portrait. There are three commonly used keys on the portrait. These are the keys do, fa, and left.



Fig. 1. The clefs are respectively “sol”, “fa” and “do” keys

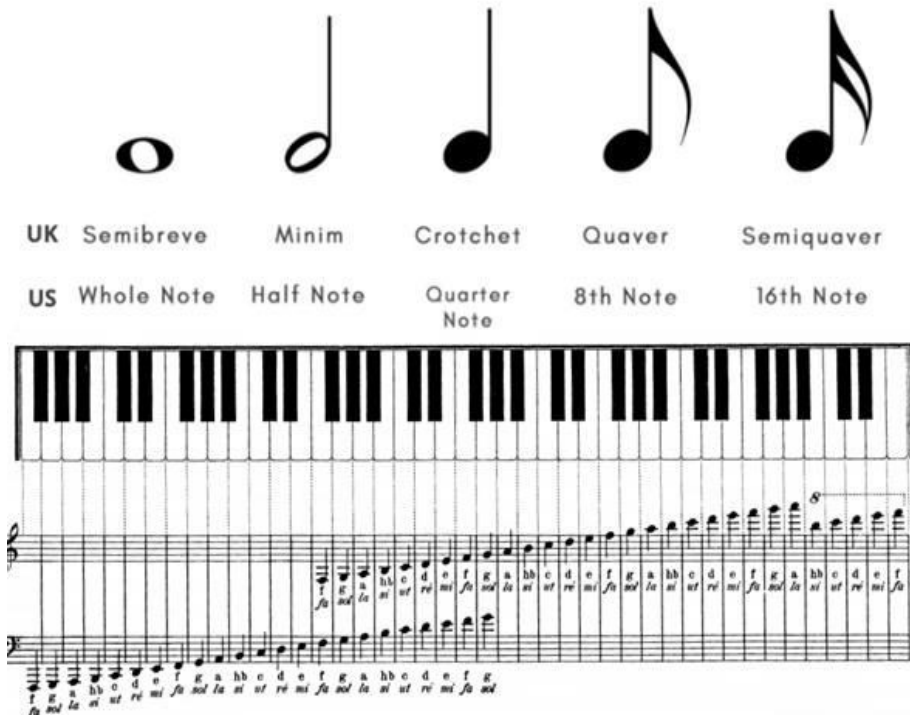


Fig. 2. Note values and positions

“Phone application that reads and plays notes with image processing techniques” study; It will enable people who have learned to play an instrument or who can play the instrument, when they do not have the opportunity to listen to the piece, to hear the piece they want to play in advance by processing it with image processing techniques on mobile devices. Hearing the piece beforehand greatly shortens the process until it can analyze it and play it. Along with the developing technology, in some sectors, it is aimed to help people who have learned or are learning an instrument, since it can be done by phones whose functionality is increasing day by day, instead of people whose duty is to show us how the job is done and to control us and correct our mistakes. As a result of the researches, similar note recognition applications have been made in other languages such as “Python” and “C++” using ready-made libraries. However, in this project, it was developed in order to organize applications made for other purposes in Java language for this purpose and to continue to work correctly in the face of variable situations and to ensure that image processing techniques are used. A number of methods have been developed for faster implementation.

Android studio: The official integrated development environment for Android application development. It is an intelligent code editor, debugger and performance analyzer designed to develop Java language applications with high efficiency results on android devices such as mobile and wearable technologies. For these reasons, while creating our mobile application, coding was done in the Android studio application.

OpenCV: It is an image-sight-based open source library used to build real-time applications. It is used very actively by many software companies. Thanks to “Computer Vision”, it is now able to enable devices not only to record cameras, but also to read the license plates of vehicles autonomously, and detect faces and objects with its advanced technological structure. Then the identification process begins. While the classification of human actions in the videos is in question, tracking of moving objects and being able to extract 3D models are the features that are beneficial. It is also useful in the production of 3D point clouds. The OpenCV program, which also enables images to be combined in high resolution, is already used in many different fields. The OpenCV library has been preferred in the image processing part due to its open source code, ease of use and compatibility with many software platforms.

2.1. Note analysis with image processing technique

The duration and tones (frequency values) of the notes are determined using image processing techniques in the image of a musical note paper in digital format. Image files that can be identified by electronic media devices are called “digital images”. A digital image consists of small elements called "pixels", each containing a single color tone. Pixels are arranged horizontally and vertically on the coordinate plane to form the digital image. In order to make sense of the symbols on the music sheet, some operations can be applied on the symbols by scanning them in pixel size. The dimensions of these symbols in pixels vary according to the resolution of the image.

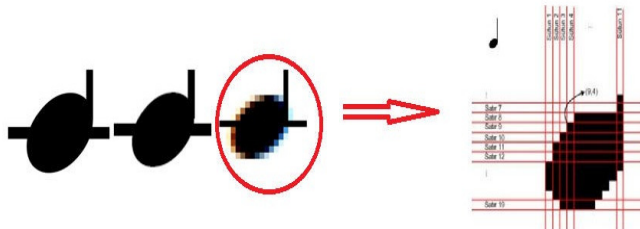


Fig. 3. Detecting the duration and tone of the notes in the image of the music sheet

Each pixel is represented by a black “0” and a white “255” within the scope of the “binary image” formed by converting to black or white. This process is not a problem, as different colors in the notes do not matter.

3. Determining the notes

In terms of image processing, the process of determining the note corresponds to obtaining information such as determining the staff, whether the note's place on the staff is full, whether the note is full, whether there is an extension point or tail next to it.

First of all, it is necessary to determine the location of the staff lines while processing the image. For this process, a binary image is obtained by equating the values above 170 to 255 and the values below to 0 on the scanned music sheet. By scanning horizontally regionally, the program is introduced to the program that there is a staff line in the areas where the mathematical average values of the pixels are below 100 and the white density is below 0.30. Finally, reference note images are searched in the photo.

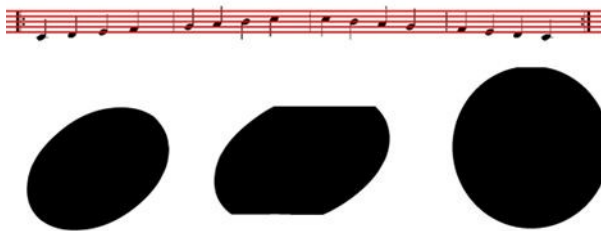


Fig. 4. The position of the note on the staff

In this method used, the reference image is searched across the entire image. The reference images shown in the previous slide were searched by this method.

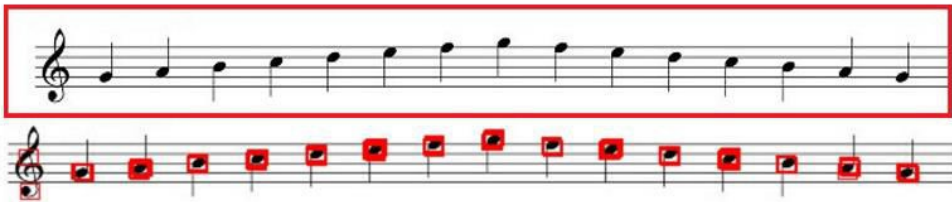


Fig. 5. Searching reference note images within the photo

4. Problems in system design

Despite the fact that the notes are shaped around the staves, in this method, since the reference image is searched in the whole photograph, it is matched both in the parts where there are no notes, and it loses time because it searches for more places than necessary. In addition, since it gives different results in the same search type, a certainty cannot be obtained.



Fig. 6. Results found in a general search on Example-3 note strings

In addition, when the similarity rate is low, it finds places close to the same place more than once. When the similarity rate increases, it cannot find the place it needs to find.

The problems we encounter in regional search are as follows. Since the quality and size information of the image changes while searching, the sizes of the shapes that allow us to

recognize the notes vary in pixels. In order to ensure a correct search, the dimensions of the image to be scanned should be the same as the size of the shape in the note image. In order to solve this problem, the problem is solved by reducing the dimensions of the quality image to be scanned to a size that corresponds to a line spacing in the note image and by reducing the width at the same rate.



Fig. 7. Results found in a general search on Example-4 note strings

5. Results

By applying etching and spreading operations, the note lines and stems of the notes in the image are recorded. Here we enlarged the ellipse parts of the notes again and a general search was performed again.

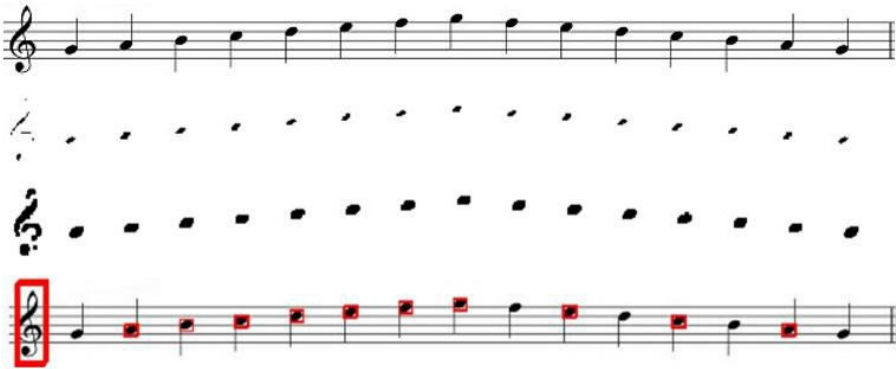


Fig. 8. Etching, spreading operations on the musical note portrait and the result found in the general search

Since some notes could not be found in the search, the similarity rate was reduced and a different image was searched. Some notes were displayed more than once, although it was possible to find all the notes more precisely (Fig. 9). For this, this problem is solved by taking the coordinate values with the difference of more than 1 and throwing them into the array (Fig. 10).



Fig. 9. Morphological and general search results



Fig. 10. As a result of applied morphological, general search and single-frame reduction operations

6. Conclusions

Notes displayed with this application will be able to be parsed and played. The program can analyze the image taken from the camera or the handwritten note image. By processing the notes it analyzes and the sound it receives from the microphone, the program is able to listen to someone playing a musical instrument and match the music to check whether it is playing correctly. The program has the ability to write the notes of the music played on its own music sheet.

You can also play different instruments with this program. It has been ensured that people learning an instrument can learn the piece more quickly by listening to the music via the phone. In addition, instrument learners can control themselves via the phone without the need for any other person. With this program, different instruments can be added and the program can evolve into a program that provides instrument training in the future.

The program was created to analyze a note sequence selected from the gallery by general search after morphological operations, and the codes were uploaded to Github. This program will be developed over time as outlined in future developments of the project in the findings and discussion section.

References

- [1] S. A. Azizi, "Brain to music to brain!," *Neuroscience Letters*, Vol. 459, No. 1, pp. 1–2, Jul. 2009, <https://doi.org/10.1016/j.neulet.2009.04.038>
- [2] M. Boşnak, A. H. Kurt, and S. Yaman, "Beynimizin müzik fizyolojisi," *Kahramanmaraş Sütçü İmam Üniversitesi Tıp Fakültesi Dergisi*, Vol. 12, No. 1, pp. 1–1, Apr. 2017, <https://doi.org/10.17517/ksutfd.296621>
- [3] Rotaache M. and Ubieta G., "Videoarte: la evolucion tecnologica. nuevos retos para la conservacion," in *8th Conference of Contemporary Art*, 2007.
- [4] Akengin G. and Arslan A. A., "Digital installation as an element of expression in contemporary art," (in Turkish), *Journal of Art Education*, Vol. 9, No. 2, 2021.
- [5] H. Mamur, A. Aktaş, and S. Kuzey, "Müzik Notalarının Matlab Yazılımıyla Belirlenmesi," (in Turkish), *Bilge International Journal of Science and Technology Research*, Vol. 2, pp. 109–115, Dec. 2018, <https://doi.org/10.30516/bilgesci.491148>
- [6] Wei-Chen Lee et al., "The realization of a music reading and singing two-wheeled robot," in *IEEE Workshop on Advanced Robotics and Its Social Impacts*, 2007.
- [7] R. Waranusast, A. Bang-Ngoen, and J. Thipakorn, "Interactive tangible user interface for music learning," in *2013 28th International Conference of Image and Vision Computing New Zealand (IVCNZ)*, pp. 400–405, Nov. 2013, <https://doi.org/10.1109/ivcnz.2013.6727048>
- [8] Y. Zhang, "Adaptive multimodal music learning via interactive-haptic instrument," in *New Interfaces for Musical Expression*, 2019.
- [9] U. Gürman, "The effect of the Synthesia program on student achievement in piano education," (in Turkish), Afyon Kocatepe University, 2019.