Computer Vision for Music

Shikhar Saikia
Computer Science and Engineering Dept.
Thapar Institute of Engineering and Technology

Abstract

There have been many endeavours on the road to making computers understand and interpret objects present in a visual format. Studies in this field have led to this interpretation of an Optical Music Recognition system which is capable of interpreting and digitizing sheet music.

1. Introduction

Culturally speaking, there are only a few mediums that have been a part of the lives of people from all over the globe for generations. One of these mediums, is that of Music. It has been a part of human culture for centuries and the origins of music are thought to have come about simultaneously or even before language.

With the growth in the last century, we have seen the use of technology dictate the way through which we consume music. It evolved from the more physical form of Vinyl's, cassettes, and CDs to the more digitized form of streaming platforms. While major jumps have been made in the production and marketing of music as a product, not much has been done for converting music from its documented form to a digitized audible version of it. The focus of this project is to achieve just that and convert notated sheet music to its auditory equivalent through the help of OpenCV.

Commented [SS1]:

2. Related Word

Optical music recognition (OMR) is a field of research that investigates how to computationally read musical notation in documents. The goal of OMR is to teach a computer how to read and interpret sheet music and produce a machine-readable version of the written music score. Once captured digitally, the music can be saved in commonly used file formats like MIDI for playback.

Early research in OMR was conducted by Ichiro Fujinara, Nicholas Carter, David Bainbridge, and Time Bell. These researchers developed many of the techniques that are still being used today.

The first commercial OMR application, MIDISCAN was released in 1991 by Musitek Corporation.

3. Method and Models

3.1 Approach

Music notation comprises a number of characters and alphabet which is used to interpret and communicate ideas. Each alphabet and character creates a visual manifestation of the note which includes its pitch and duration. Music notes are presented on a staff, a set of 5 lines each having its own pitch. The pitch of a given note is determined by its vertical position on the staff and the duration of a given note is determined by its horizontal position on the staff. Thus, music notation presents information in two dimensions. There are a countless number of alphabets and symbols in music notation and thus creating a computer vision system that can interpret the same is a tedious but achievable goal.

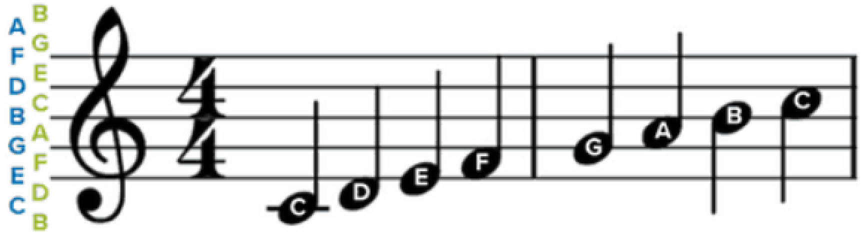

Fig 1. Example of how a general music score looks

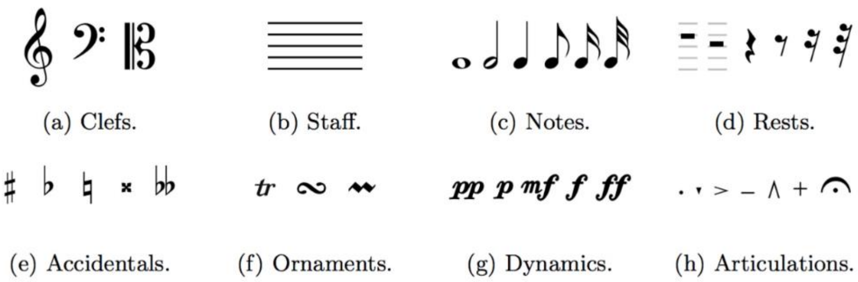

Fig 2. Types of symbols in a music score

3.2 Scope

In this project, I seek to create an OMR system which should be capable of recognizing sheet music through images. To not over complicate the process, I will be utilizing simple sheet music written for a single instrument. This system will not be able to detect tempo, tuplets, repeats, or dotted rhythms.
It aims to simply be able to read and convert uncomplicated melodic compositions into a playable format.

3.3 Framework

The general framework of an OMR system can be summarized as:

1. Image Processing
2. Music Primitive Detection
3. Music Primitive Classification
4. Semantic Reconstruction

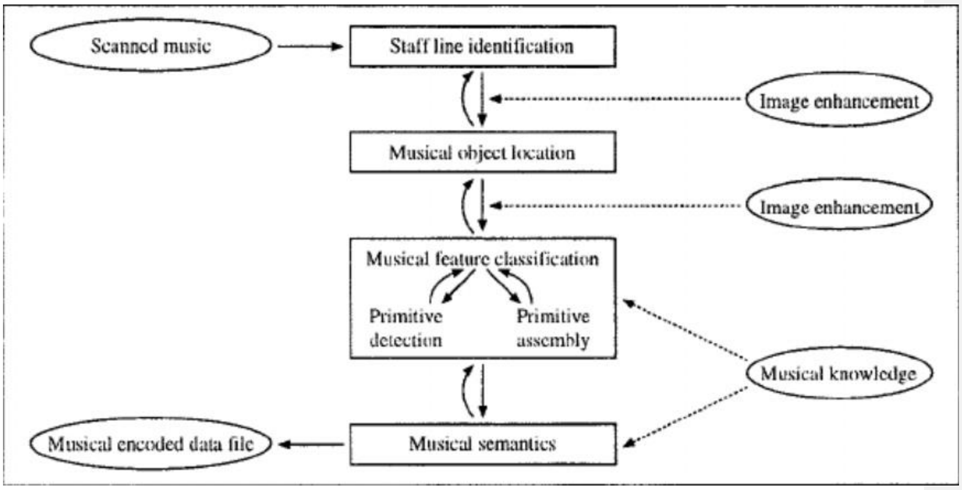

Fig 3. Framework of an Optical Music Recognition (OMR) System

3.4 Pre-processing

The main task of this stage in this project is to make changes to the original music document such that the recognition process becomes easier and more efficient. This involves image enhancement, noise removal, etc.

Only a few pre-processing operations are necessary, the system performs binarization and noise removal. Noise removal attempts to remove image artifacts if present and binarization converts the image into a matrix made of binary values.

Staff Detection

As mentioned before, Music notes are presented on what is called a staff. It is a coordinate system ranging in two dimensions. Owing to its importance, finding the staff lines is the most important initial step in the detection of the score. *Staff line thickness* and *Staff spacing height* are calculated using Run-Length Encoding (RLE). Wherever the colour black is most common, the *staff line thickness* is approximated and wherever the colour white is most common, the *staff space height* is approximated.

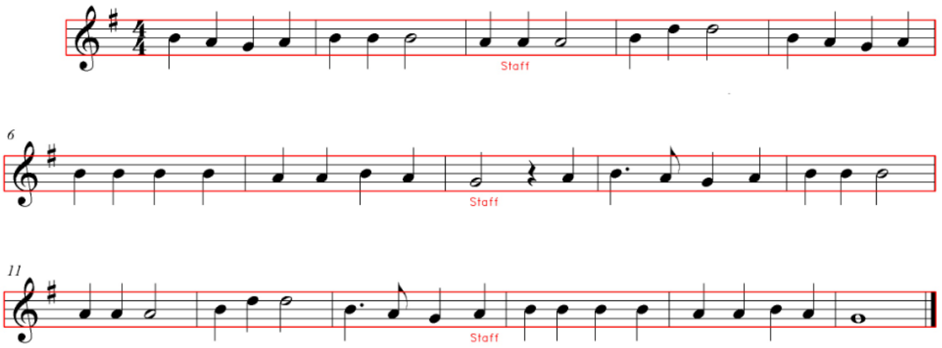

Fig 4. Staff Detection

Music Primitive Detection and Classification

- **Music Primitive Detection:**

After the process of Staff Detection is complete, the primitives present in the score are detected by comparing the obtained score image to a template of the same. For this to work smoothly, I have created a library of labelled primitives which houses various images of notes, clefs, barlines and staffs to make the process more efficient.

- **Music Primitive Classification**

Once the process of Music Primitive Detection is complete, the various notes, alphabets and clefs detected are classified staff wise into their respective categories.

Fig. 5 Music Primitive Detection and Classification

Semantic Reconstruction

After all of the primitives have been detected and classified, they are sorted in their correct order on the staff from left to right. Once the primitives have been reconstructed on each staff in the score, all the primitive values are then compared with a number corresponding to its associated MIDI note value and added to a MIDI track.

Result and Discussion

The performance of this system can be assessed in both a qualitative and a quantitative way. As the outcome so produced at the end of the system is in an audible medium, one can listen to the file and pinpoint its accuracy to the true notes written on the score. This, although a good way to arrive to conclusion, is a time consuming process and cannot be accomplished without decent prior knowledge of musical theory and instruments. Therefore, a quantitative assessment of this system would provide a better result.

Overall, the system performed well in terms of detecting the different kinds of notes and did manage to produce a playable output which was quite accurate. It did however have issues with the detection of time signatures and clefs at occasion. It seemed to work well with uncomplicated compositions and did somewhat struggle with bigger pieces.

Conclusion

Although working with musical notations has been quite fascinating, I have realised it is quite difficult to create an OMR system which is flexible and can handle complicated compositions. The problem arises due to the multitude of acceptable symbols which can be implemented while writing music. A machine learning approach in this scenario may help in better classification and reading of music scores. Time is also a factor which may be majorly improved as it currently requires a large span of time to generate the MIDI file.

References

1) Music Notation and Theory for Intelligent Beginners by Jono Kornfeld
2) Understanding Optical Music Recognition ~Jorge Calvo-Zaragoza, Jan Hajic Jr., Tu Wien, Alexander Pacha (May 2020)
3) State of the art Model for Music Object Recognition with Deep Learning ~ Zhiqing Huang, Xiang Jia, Yifan Guo (June 2019)
4) Optical Music Recognition: State of the Art and Major Challenges ~ Elona Shatri, Gyorgy Fazekas, University of London (June 2020)