RECOGNIZING COMPOSER'S MUSICAL SIGNATURE FROM BARS OF MUSIC USING COMPUTER VISION AND DEEP LEARNING

By

Ashim Shome

ID: 183-25-718

Department of Computer Science and Engineering

Daffodil International University

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Science and Engineering

Supervised By

Ahmed Al Marouf

Lecturer

Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

DECEMBER 2019

APPROVAL

This Thesis titled "**Recognizing Composer's Musical Signature from Bars of Music using Computer Vision and Deep Learning**", submitted by Ashim Shome, ID No: 183-25-718 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 6th December 2019.

BOARD OF EXAMINERS

____

Dr. Syed Akhter Hossain **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

____

Dr. Md. Ismail Jabiullah **Internal Examiner**
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

____

Dr. Sheak Rashed Haider Noori **Internal Examiner**
Associate Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

____

Dr. Mohammad Shorif Uddin **External Examiner**
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

I hereby declare that, this thesis has been done by me under the supervision of **Ahmed Al Marouf,** Department of CSE, Daffodil International University. I also declare that neither this thesis nor any part of this thesis has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Ahmed Al Marouf
Lecturer
Department of Computer Science and Engineering
Daffodil International University

Submitted by:

Ashim Shome
ID: 183-25-718
Department of Computer Science and Engineering
Daffodil International University

ACKNOWLEDGEMENT

First, I express my heartiest thanks and gratefulness to almighty Allah (God) for His divine blessing makes me possible to complete the M.Sc. thesis successfully.

I really grateful and wish my profound my indebtedness to **Ahmed Al Marouf, Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of "*Computer Vision*" to carry out this thesis. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this thesis.

I would like to express my heartiest gratitude to **Dr. Syed Akhter Hossain, Professor and Head**, Department of CSE, for his kind help to finish my thesis and also to other faculty member and the staff of CSE department of Daffodil International University.

I would like to thank my entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, I must acknowledge with due respect the constant support and patients of my parents.

ABSTRACT

For centuries, music has been divided over two traditions in the form of written documents and aural transmission normally called musical scores. Many of these scores are published in manuscript form and So they are at risk of being lost over time. The system takes a music score image as input, segments music symbols after preprocessing the image, then recognizes their pitch and duration. Finally, MIDI files are generated. Similar to optical music recognition (OMR) systems, programs similar to optical character recognition systems have been in intensive development for many years. This thesis provides an overview of automatic analysis of handwritten music scores. An overview of the literature provides an overview of OMR processing systems for the benefit of the reader and self-interest. The OMR system can provide various benefits to the scientific community for an effective and powerful printed and handwritten music score. We have presented some of the strategies of OMR utilizing computer vision and deep learning based approaches for music scores and printed processing.

TABLE OF CONTENTS

CONTENTS	PAGE

LIST OF FIGURES

LIST OF TABLES

Chapter 1
Introduction

1.1 Introduction

Transcript of sheet music can be carried out manually in a machine readable format. However, the complexity of the music transcription leads to heavy software for resolution, for this heavy software the whole process very prone to errors. So Optical Music Recognition (OMR) system, Which is recommended, for automatic transcription of music score image. As input the OMR system accepts the music score image and automatically converts it to some symbolic structure such as MIDI [1].

An OMR system is created of three sub-modules: symbol segmentation, image preprocessing, and symbol recognition. Most of the research focuses on recent staff line identification, staff line removal, music symbol segmentation and note recognition. The most important thing is to removing the staff's lines. It simplifies the process of segmentation, recognition and representation symbols. for applied to detect staff lines Stable path method [2][3] and Hough Transform [4]. The another difficult problem is without destroying the music notes how to remove staff lines. Chen and Xia [5] is a morphological operation for music score image process. Detecting the rectangular note boundaries then The staff lines were removed. However, the operation is more or less corrupted by symbols on the staff line, So VO [6] and Mehta [7] do processes without removing the stave lines.

The task of the music symbols segmentation after removing the staff line. The another important part of Segmentation in the entire process, If the music notes are properly segmented and if they can be recognized correctly. Hierarchical digestion [8] is the most common approach for symbol segmentation. First analyzed and separated by staff lines, and then the primary graphic symbols were extracted from a music sheet. To segment the symbols as well Template matching [9] and projection method [6][10] were applied.

all the basic symbols segmented by Wen [11]. Mehta [6] and Blanes [12] offer to keep the notes head and stem flag connected. After segmented the symbols are identified as specific duration and pitches. For note identification Minimum Spanning Tree Algorithm [15] is also a popular method

Recently, for recognize music notes neural networks, Recurrent Neural Networks [1] and Convolutional Neural Network [12][14].

The proposed system can convert printed music scores to MIDI files. Firstly the staff lines are remove using Morphological operations [5] and for segment the symbols projection method [6] is used. Then neural networks [14] used for music notes recognizing. The last stage is MIDI transformation.

1.2 Motivation

In every sense of happiness, sorrow and sorrow of life, the melody of the tune is constantly sounding in the secret place of the heart. People who like to listen to more and less, do not get caught up in the love of music - it is very difficult to find people on earth. You need to overcome the nervousness and tiredness of your life. It can touch our feelings and it can also change them.

Human Brain Can Easily Understand and read musical notes. But the human brain takes a lot of effort and effort to understand and read its notes. In this way the computer cannot easily understand and read these musical notes.

In this research to help the computer understand how these musical notes are easy to understand and readable. My task is to help the computer more easily know how to make these notes easy to respected outcome. The goal of my thesis paper is to create proficient workflow management system to minimize the price and complexity of changeable large, save it to digital form.

During the late 1980s the flat-bed low-cost digitizers skill to a prolongation of OMR research activities. Many are commercially available software but no software has been created to robustness satisfy them [6] even the most advanced acceptance products cannot recognize all musical symbols until now. Handwritten music cannot bring good results with printed documents.

1.3 Rationale of the Study

There are many optical music recognition system algorithms are available in the market and Researchers are trying to build new algorithms. Many are commercially available software but no software has been created to robustness satisfy them [6]. even the most advanced acceptance products cannot recognize all musical symbols until now. Handwritten music cannot bring good results with printed documents. This type of research is very helpful for the Deep Learning and improve the fields of Computer Vision, Computational modeling. The ability of a music symbol to detect music is noteworthy, although it is very difficult for a machine to identify the symbol individually.

People can learn thousands of faces throughout their lives and even recognize familiar faces at a glance, even after years of isolation [32]. promises for Optical Music Recognition (OMR) to make large-scale collections of sheet music searchable by their algorithms. It will describe fancy ways of accessing wide quantity of written music. As promised for a long time OMR wasn't doing that, because its ability is not good enough. Especially in terms of non-standard images or handwritten songs. even so, for advances in machine learning. We're talking about the study of music education that can Transcription at very low cost the OMR outputs.

1.4 Research Questions

RQ1: Can musical signatures be recognized using computer vision based approaches?

RQ2: How to improve the recognition of a composer’s musical signature using computer vision?

1.5 Expected Output

The proposed system can convert printed music scores to MIDI files. Firstly the staff lines are remove using Morphological operations [5] and for segment the symbols projection method [6] is used. Then neural networks [14] used for music notes recognizing. The last stage is MIDI transformation.

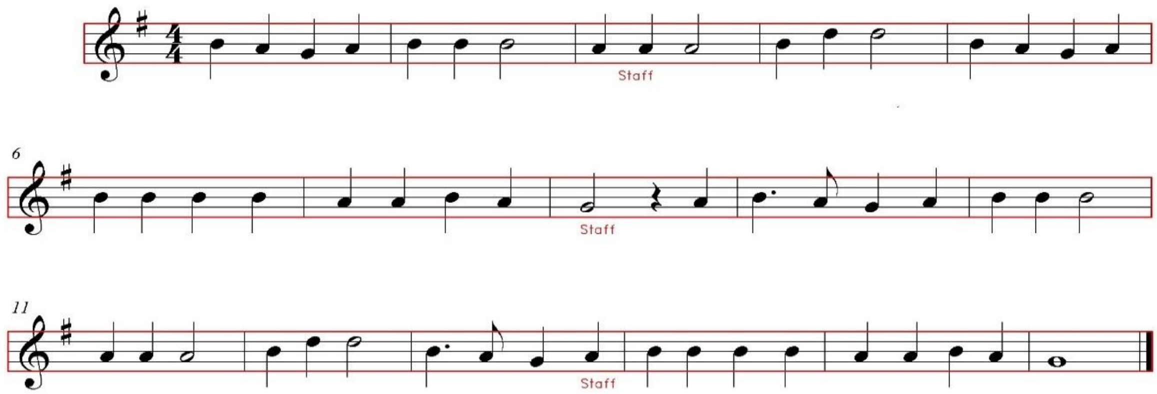


Figure. 1.1 Input music notes

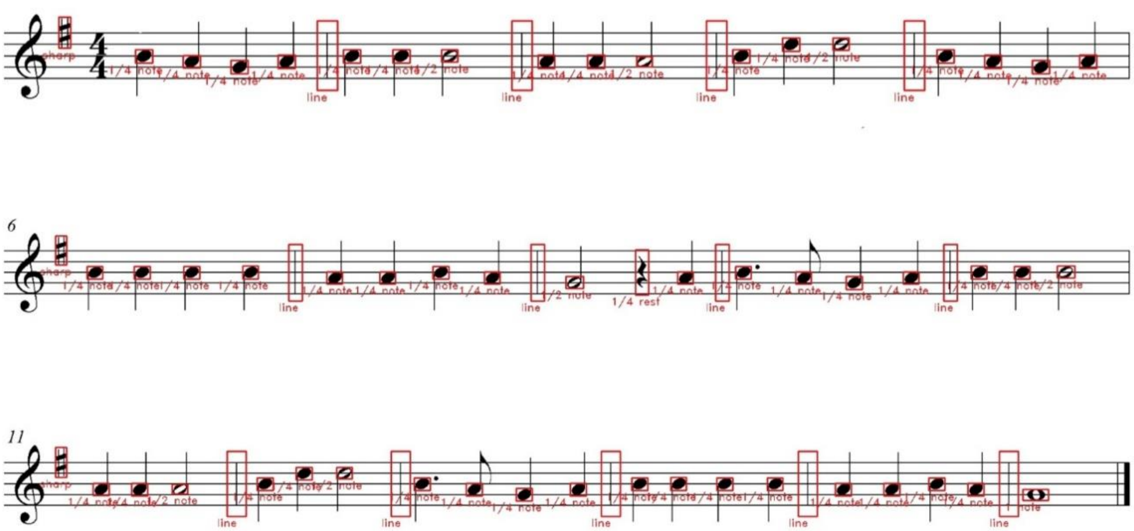


Figure 1.2. Output music notes with tagging.

CHAPTER 2

Background

2.1 Introduction

The computer cannot easily understand and read these musical notes. In this research to help the computer understand how these musical notes are easy to understand and readable. My task is to help the computer more easily know how to make these notes easy to respected outcome. The goal of my thesis paper is to create proficient workflow management system to minimize the price and complexity of changeable large, save it to digital form.

During the late 1980s the flat-bed low-cost digitizers skill to a prolongation of OMR research activities. Many are commercially available software but no software has been created to robustness satisfy them [6] even the most advanced acceptance products cannot recognize all musical symbols until now. Handwritten music cannot bring good results with printed documents.

2.2 Related works

As ever growing and an interesting topic Recognizing a composer's 'musical signature' from a bars of music using Computer Vision, attracted many researchers. Some of the listed and worth mentioning are briefly described below.

Table 2.1 Best Works Fit Review

S.no	**Title**	**Year**	**Author Name**	**Journal/Conference**	**University/ Affiliation**
1	Old Handwritten Music Symbol Recognition using the Combination	2018	Savitri Apparao Nawade, Rajmohan Pardeshi, Chitra Dhawale, Mallikarjun Hangarge	International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU).	P.R.Pote college of Engineering Amaravati India and Karnatak Arts Science and

	of Foreground and Background Projection Profiles				Commerece College, Bidar,India
2	Camera-Based Optical Music Recognition Using a Convolutional Neural Network	2017	Adri`a Rico, Alicia Forn´es	International Conference on Document Analysis and Recognition (ICDAR)	Universitat Autonoma de Barcelona Bellaterra, Catalonia, Spain
3	Optical Music Notes Recognition for Printed Music Score	2018	Chuanzhen Li, Jiaqi Zhao, Juanjuan Cai, Hui Wang*, Huaichang Du	International Symposium on Computational Intelligence and Design (ISCID).	Communication University of China Beijing, China
4	An optical music recognition system for the notation of the Orthodox Hellenic Byzantine Music	2015	Velissarios G. Gezerlis, Sergios Theodoridis	International Conference on Pattern Recognition. ICPR	University of Athens, 157 84, Athens, Greece
5	An	2016	Fu-Hai Frank Wu	IEEE International	National Tsing

	Evaluation Framework of Optical Music Recognition in Numbered Music Notation			Symposium on Multimedia (ISM)	Hua University, Hsinchu, Taiwan
6	Pen-Based Music Document Transcription	2017	Javier Sober-Mira,1Jorge Calvo-Zaragoza,2David Rizo,1and Jos´eM.Ĩnesta1	International Conference on Document Analysis and Recognition (ICDAR)	University of Alicante, Alicante, Spain and McGill University, Montr´eal, Canada
7	Optical music recognition for traditional Thai sheet music	2014	Worapan Kusakunniran1, Attapol Prempanichnukul2, Arthid Maneesutham3,	International Computer Science and Engineering Conference (ICSEC)	Mahidol University, Thailand

2.3 Research Summary

For MIDI most music symbol systems import and export the final representation of a musical score. However, over the years different more music encoding formats has been created for the music. The OMR systems used are compliant andthrough the use of non-adaptive stars do not improve their performance. However, it is now a challenge. Furthermore only recognition of the highest OMR system for printed music scores result. This often fails when the input image is in the form of poorly documented photocopies or documents, only during photo scores that work with exception handwriting scores.

Table 2.2 Review The most relevant OMR programs and software

S.no	Software and program	Output file
1	PhotoScore	MIDI, WAVE,MusicXML, PhotoScore, NIFF
2	Capella-Scan	Capella, MusicXML, MIDI
3	SmartScore	Finale,NIFF, PDF, MIDI
4	ScoreMaker	MusicXML
5	SharpEye	MIDI, NIFF, MusicXML
6	Gamera	XML files
7	Vivaldi Scan	XML, Vivaldi, MIDI
8	Audiveris	MusicXML

2.3 Basic music notes names

Table 2.3 Basic music notes names

S.no	Written notes	American time name
1		Whole note
2		Half note

3		Quarter note
4		Eighth not
5		Eighth not
6		Dotted half note
7		Sixteenth note
8		Quarter note rest
9		Half note rest

10		Whole note rest
11		Eighth not rest

2.4 Binarization

To convert a pixel image to a binary image,the name of the process is Binarization. Almost all OMR systems are started using the binarization process. It's the hob to analyze which determines which digital image is useful and which is not effective.

In the past many algorithms have been proposed. To make binarization an automatic process, the success rate varies depending on the problem. The advantage of doing OMR is that there is a great deal of degradation that reduces the amount of OMR required in the process. This facilitates higher quality efficiency and possible model design of the OMR task.

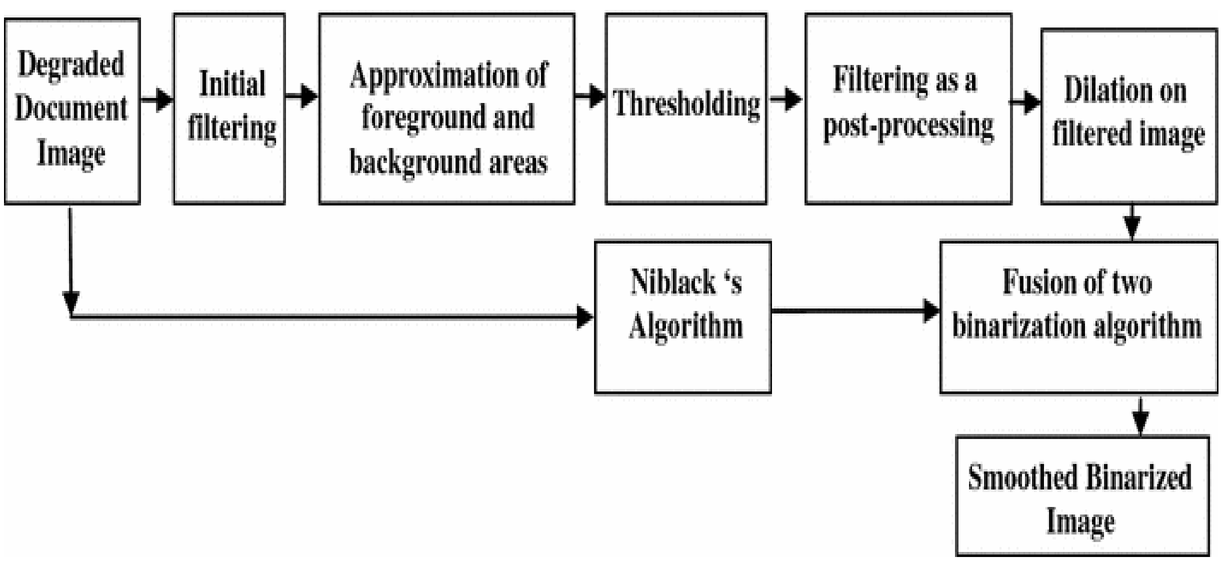
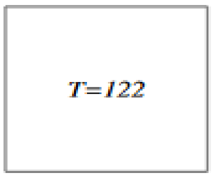


Figure 2.1 Efficient binarization technique for severely degraded

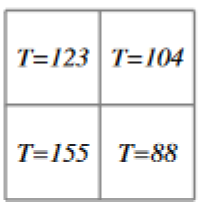
2.5 Image preprocessing

The underlying structure performed satisfactorily on test images in subjectivities. A brief description of its performance will be given below. In general The tests were based on the following objectives: As a basic criterion, the number of staff identified will be measured. Then, note length and note recognition counter-structure will be examined for greater scope of vertical objects



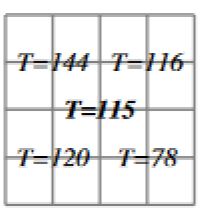
$$T_{global}(g) = \begin{cases} 0 \ldots g < t \\ 1 \ldots g \geq t \end{cases}$$

Figure 2.2 Global thresholding.



$$T_{local}(x,y) = \begin{cases} 0 \ldots g(x,y) < t \\ 1 \ldots g(x,y) \geq t \end{cases} \quad \in \text{Region } R_i$$

Figure 2.3 Local thresholding.



$$T_{adaptive}(x,y) = \begin{cases} 0 \ldots g(x,y) < t(N(x,y)) \\ 1 \ldots g(x,y) \geq t(N(x,y)) \end{cases}$$

Figure 2.4 Adaptive thresholding.

. In conclusion, the classification of other marks present in the worker system will be calculated, and the other focus will be on the classification of the pitch classification rate.

The numbers will be surveyed against Ground Truth data, in all of the cases mentioned above, which were manually com-putted. It has previously been noted that the data set contains 15 note sheets, but a specific piece was found to be inappropriate, as not the vast majority of children’s tunes, this special score was written for two voices.

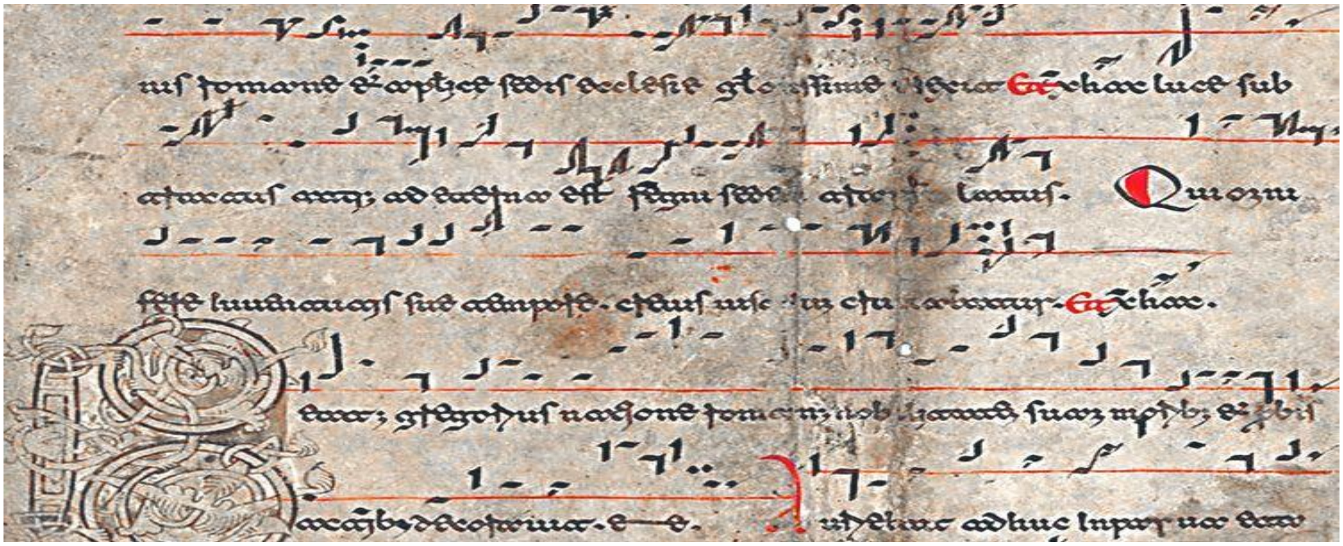


Figure 2.5 Early music notation

Figure 2.6 Modern music notation

Chapter 3

Research Methodology

3.1 Introduction

The arduous task of (OMR) has been widely studied over the past several decades and has provided numerous publications related to the field of image processing, graph theory, pattern recognition or possible coding for naming. Numerous OMR systems have become non-commercial and commercial as a result of efforts, especially from the sixties when hardware prices dropped at a fair rate. Nevertheless, OMR applications are considered as efficient compared to their OCR partners [42]. The OMR problem is usually divided into several sections, which are usually processed step by step Due to the complexity. According to Bainbridge [5] The problem of OMR is described in four main sections.Due to a different approach to distortion, an alternative terminology is proposed [34]. Furthermore, each of the above phases is subject to change, in large or small quantities, in the manner of use - these will be discussed in more detail below.

3.2 Research Subject and Instrumentation

Research subject: Computer Vision and Deep Learning.

Research Instrumentation:

Python 3.6

python-numpy

Python matplotlib

python-opencv

Python MIDIUtil

Computer (4 GB RAM, Core i5 3.2 GHz)

3.3 Proposed System Algorithm

3.3.1Algorithm 1: By projection Segmentation of symbols

Input: after preprocessing printed music score image

Output: Music symbols

1: **For** represent the number of black pixels in the image Create a empty lists 1

2: **for** I do in row

3: Into the list calculate of each row the black pixel values, and compare the pixel maximum: gMax

4: **if** The element m in the list $<$gMax × 0.95

5: horizontally Segmented imagefrom here

6: **end if**

7: **end for**

8: Create a empty lists 2 to represent the number of black pixels

9: **for** in column j do

10: Into the list calculate the black pixel values of each column and compare the pixel maximum: mMax

11: **if** The element n in the list $<$ 0.95 ×mMax

12: vertically segmented image from here

13: **end if**

14: **end for**

3.3.2Algorithm 2: Recognition pitch

Output: Pitch of music notes

1: To calculate the staff space and detect the position of the line

2: Use the template matching algorithm to detect the position of the note heads Use the Hough transform

3: The distance between the position of staff line and the note heads Determine the pitch base

Input: Printed image for music score

4: **end**

3.4 Reference lengths

Most OMR algorithms in the presence of binary images rely on an estimate that separates two staff lines drawn with staff line thickness and distance. Based on these values can be further processed and may differ from some magic numbers of predetermined. In other fields, such as fixed marginal numbers, systems are complex and adapting to new and unexpected situations makes them even more complicated.

Figure 3.1 Featured pages from Cardoso and Rebello are Stafeline_Height and StaffSpace_Height.

(a) Original music score

(b) Score binarized with Otsu's method.

Figure 3.2 Original music score & Score binarized with Otsus method

Here is one and zero input used for the recognition process. The only length required here is in this case, the RLC is more compact, for example, {1 0 1 1 0 1 1 0 1 1 0 1 1 0 0 1 1 1 1 1 1 0 0 11 1 1 0 1} can be coded as sequences 1 2, 1, 3, 2, 4, 4, 1, 5, 2, 6, 1 will start the classification (the length of zero will be used if the sequence starts with 0). By using RLEs encoding each column of the digitized score, the most common white run represents staff space_height and the most common black run represents Stateline_ height.

Nevertheless, the position of the score collected with the higher words, it can not only be identified as a list of original papers, and digitalization and binary time remain part of the changing woodwork. Impairing the quality of subsequent activities makes the results unsatisfactory1 and staff space_content = 1 during binarization (the true values are Stateline height = 5 and staff space white = 19).Cardoso and Rebello proposed the work [17], which encouraged the proposed work of [72], finding the most common sum of the two vertical runs drawn represents a more robust estimate of

3.5 Statistical Analysis

3.5.1 Staff line detection

The basic stage in many OMR systems is the identification and removal of staff lines. The reason for detecting and deleting staff lines depends on the need to separate the music symbols for more accurate and efficient identification of each symbol present in the score. There are many authors [5, 7, 45, 58, 68, 76, 93] who have suggested algorithms without removing staff lines. Here, the decision is simplified to facilitate the following tasks with the risk of hearing the noise.

.Each other the staff lines are often not horizontal or straight and in some cases rarely parallel . Furthermore, most of these works are outdated, Fujinaga [4] includes a set of image processing algorithms with run-length coding (RLC), coupled-component analysis, and estimation. To determine the thickness of the staff line and the space between the staff line, after applying the RLC, any vertical black run greater than twice the staff line will be removed.

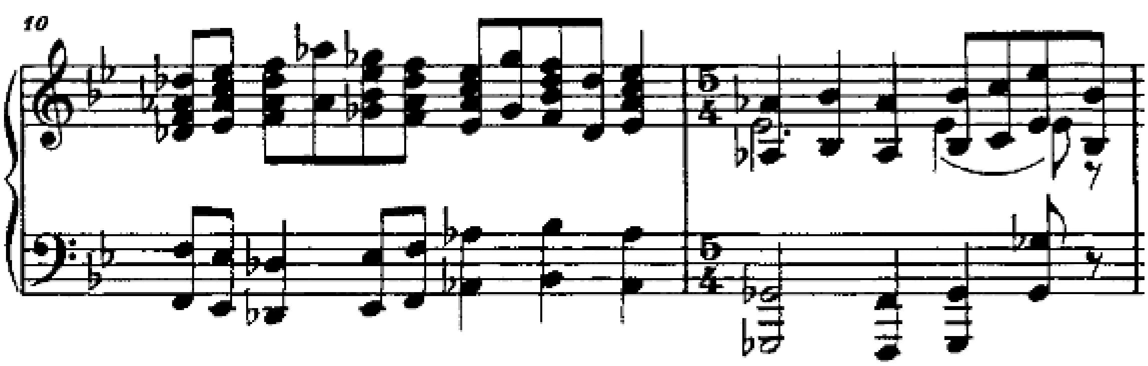


Figure 3.4 Staff line detection

The staff line has some limitations all over again Despite the variety of methods available for staff line detection. Usually some curved or paused lines are inadequately resolved. Dash detection

3.5.2 Staffline removal

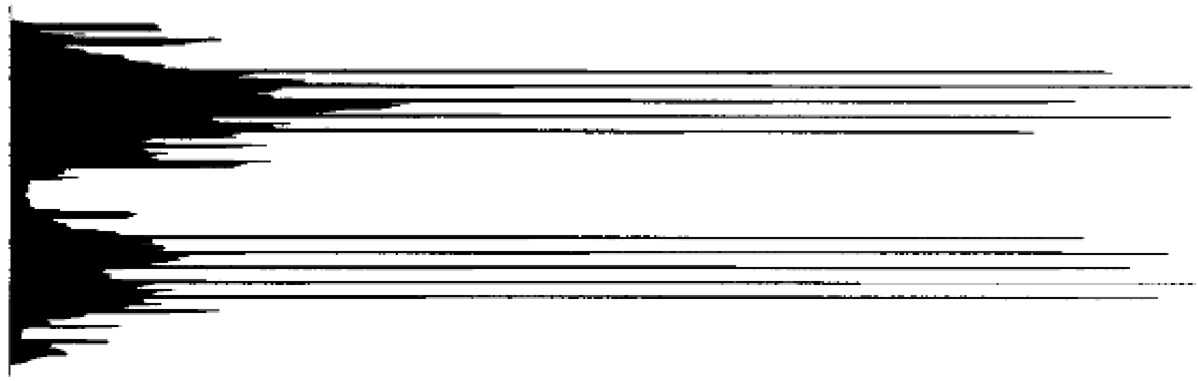


Figure 3.5 Staff line removal

The algorithm, which searches image by pixel by pixel, finds black pixel areas categorized as spots or dashes, then tries to combine dashes to create a line. A common problem with all of the above strategies, without properly incorporating global information into the detection process, is that they try to create staff lines from local information. The underlying features of the staff lines, which is true given that they are the only expansive musical score in the black object. In general, some of the most interesting strategies emerge when defining the detection process as the result of some global performance optimization. In [16] The staff line is the result of a global optimization problem where the authors propose a graph-theoretic framework.

3.6 Implementation Requirements

During the late 1980s the flat-bed low-cost digitizers skill to a prolongation of OMR research activities. Many are commercially available software but no software has been created to robustness satisfy them [6] even the most advanced acceptance products cannot recognize all musical symbols until now.

Handwritten music cannot bring good results with printed documents. This requires a typesetting or computer system that can ideally store music that can create new scores and decode the symbolic images automatically. Computer music typesetting software was developed in the 1970s and 1980s but the software was bad to use. It introduces musical keyboard data entry with slow-moving mouse and keyboard solves only a little. Still many parts and scores remain handwritten.

There is a right and strong demand for optical music recognition (OMR) system.Digitization has been generally used for Conservation as a possible tool, duplications easy, duplications simple, distribution, and the digital processing. but for facilitate operations a machine-readable format needed for example retrieval, search, and analysis. The OMR research field began with Prerau [73] and Pruslin [75] and Since then has made many important Progress. Several summaries and surveys have been. lesson in music imaging included recognition, digitalization, and restoration and also provided a well-Details list of software and hardware in OMR systems. During the late 1980s the flat-bed low-cost digitizers skill to a prolongation of OMR research activities. Many are commercially available software but no software has been created to robustness satisfy them

Similar to optical music recognition (OMR) systems, programs similar to optical character recognition systems have been development in many years. We processed the state-of-art algorithms for this music scores. This thesis provides an overview of automatic analysis of handwritten music scores. An overview of the literature provides an overview of OMR processing systems for the benefit of the reader and self-interest. The OMR system can provide various benefits to the scientific community for an effective and powerful printed and handwritten music score. In this article, we present some of the strategies of OMR We can use it in music scores and printed processing.

Chapter 4

Experimental Results and Discussion

4.1 Introduction

Digitization has been generally used for Conservation as a possible tool, duplications easy, duplications simple, distribution, and the digital processing. but for facilitate operations a machine-readable format needed for example retrieval, search, and analysis. The OMR research field began with Prerau [73] and Pruslin [75] and Since then has made many important Progress. Several summaries and surveys have been. lesson in music imaging included recognition, digitalization, and restoration and also provided a well-Details list of software and hardware in OMR systems.

4.2 Experimental Results

In this thesis there are many handwritten symbols. In the paper only printed music symbols are used. We put every note symbol on a large enough block. We have resized the row vector to the input size in neural networks. We use two general neural network methods. The segmented symbols derived the test data by Pitch Recognition. We obtain about five hundred symbols by segment some common music score images. After twelve iterations, using a three-layer convolutional neural network, we complete this process. The results demonstrate CNN's ability for image classification. In the thesis, we simply compare the results of note recognition to the printed music score. We segment some common music score images and obtain about 500 symbols. Using a 3-layer Convolutional Neural Network, after 12 iterations, the recognition accuracy of 98.47% has been achieved. Using a 3-layer LSTM, after 25 iterations, the recognition accuracy of 94.64% has been achieved. The results show the efficiency of the CNN for image classification.

2/4	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0
4/4		1.0	0	0	0	0	0	0	0	0	0	0	0	0
C-Clef	0	0	1.0	0	0	0	0	0	0	0	0	0	0	0
Eight-Note	0	0	0	0.96	0	0	0	0	0	0	0.04	0	0	0
Eight-Nest	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0
F-Clop	0	0	0	0	0	1.0	0	0	0	0	0	0	0	0
G-Clop	0	0	0	0	0	0	1.0	0.11	0	0	0	0	0	0
Half-Note	0	0	0	0	0	0	0	0.98	0	0	0	0	0	0
Quartwe-Note	0	0	0	0	0	0	0	0	1.0	0	0	0	0	0
Quartwe-Rest	0	0	0	0	0	0	0	0	0	1.0	0	0	0	0
Sixteenth-Note	0	0	0	0	0	0	0	0	0	0	1.0	0	0	0
Sixteenth-Rost	0	0	0	0	0	0	0	0	0	0	0	0.9	0	0
Whole-Half-Note	0	0	0	0	0	0	0	0	0	0	0	0	1.0	0
Whole Note	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0

Figure 4.1Confusion matrix of CNN

2/4	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0
4/4		1.0	0	0	0	0	0	0	0	0	0	0	0	0
C-Clef	0	0	1.0	0	0	0	0	0	0	0	0	0	0	0
Eight-Note	0	0	0	0.51	0	0	0	0	0	0	0.19	0	0	0
Eight-Nest	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0
F-Clop	0	0	0	0	0	1.0	0	0	0	0	0	0	0	0
G-Clop	0	0	0	0	0	0	1.0	0	0	0	0	0	0	0
Half-Note	0	0	0	0	0	0	0	1.0	0	0	0	0	0	0
Quartwe-Note	0	0	0	0	0	0	0	0	1.0	0	0	0	0	0
Quartwe-Rest	0	0	0	0	0	0	0	0	0	1.0	0	0	0	0
Sixteenth-Note	0	0	0	0.45	0	0	0	0	0	0	0.57	0	0	0
Sixteenth-Rost	0	0	0	0	0	0	0	0	0	0	0	1.0	0	0
Whole-Half-Note	0	0	0	0	0	0	0	0	0	0	0	0	1.0	0
Whole-Note	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0

Figure 4.2 Confusion matrix of LSTM

In the thesis, we only compare the results of notes recognition for printed music score. The comparative results against other methods are listed in Table 4.1. The results show that our method is superior to the general methods.

Table 4.1 Comparison of our method to other reported methods

Methods	**Accuracy**
Artificial neural network	92.38%
Combined neural network	98.82%
Minimum spanning tree	97.9%
The proposed method of LSTM	94.64%
The proposed method of CNN	98.47%

4.3 Descriptive Analysis

4.3.1 Proposed work

The proposed system is completed in five steps: image preprocessing, symbol segmentation, pitch recognition, duration recognition and last MIDI transformation. Input images are gray-scaled and the image is binarized into pre-processing. Then, after moving the staff lines, the image is segmented. The third steps is pitch and duration of the symbol recognized are fourth steps, respectively. MIDI transcription is the last step.

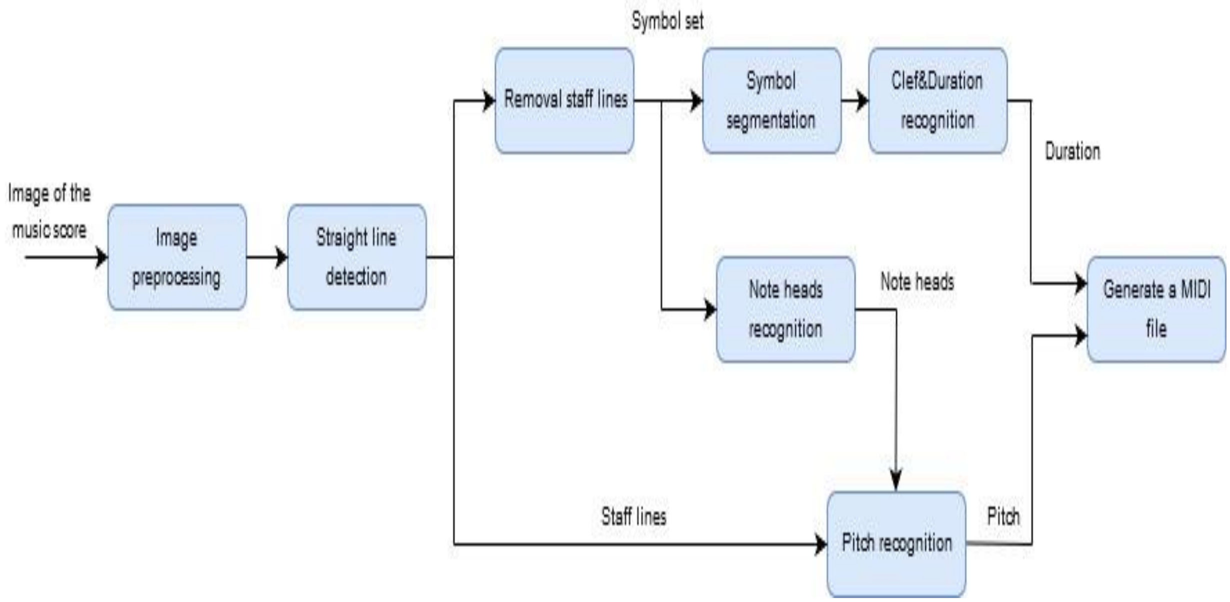


Figure. 4.3The proposed OMR system.

4.3.2 Symbol Segmentation

Firstly symbol is segmented before the staff lines are removed. To accomplish this the Morphological operations are used. The image was enlarged using rectangular structural material and was retained solely for musical symbols. After the lines of staff are removed, the music score sheet will be left only with the interest of the symbols. Thus, the process of symbolism, division and representation will be much smoother.

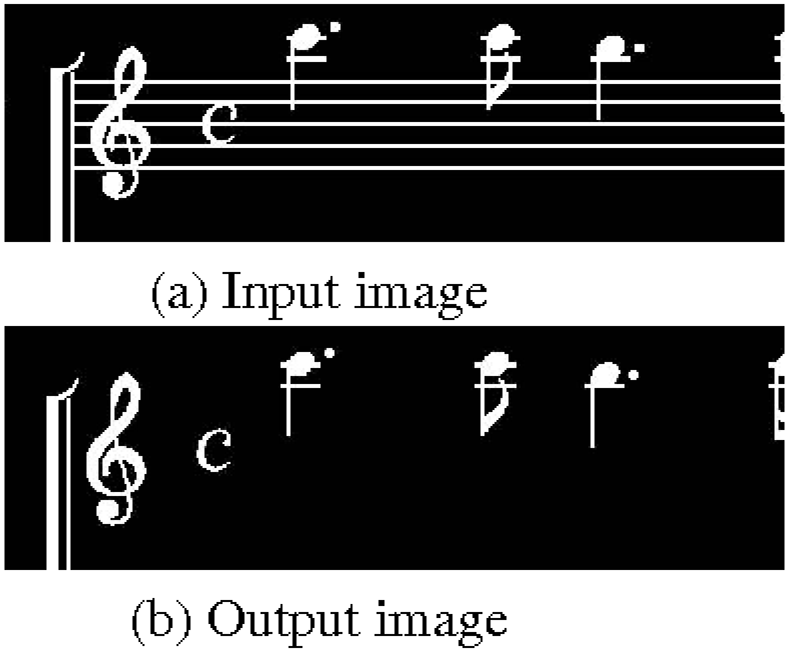


Figure 4.4The image of the music line removing

First, to the histogram obtain, the image is horizontally computed with a black pixel, which is then sectioned at the trough position. The resulting black pixel of the result also gives a computed histogram of the image's vertical projection and is divided into trough positions. At the last, single musical symbol is received. The overall algorithm as shown in summarizes.

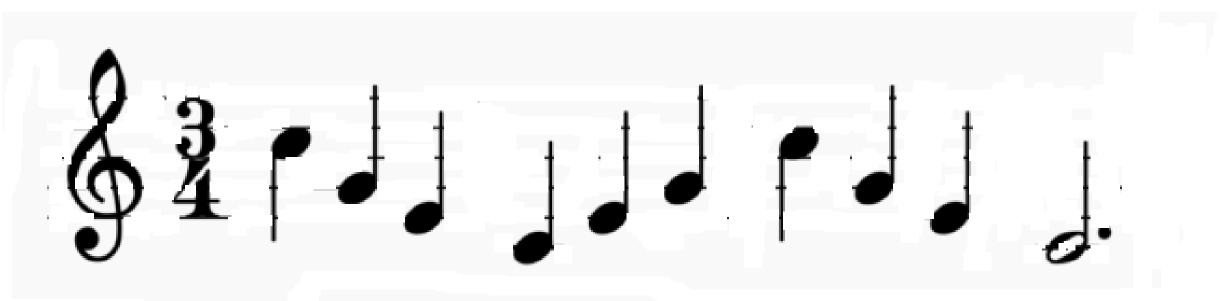
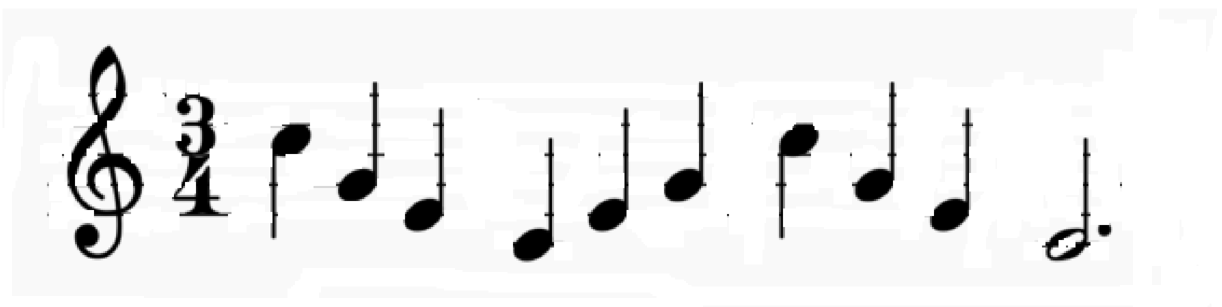


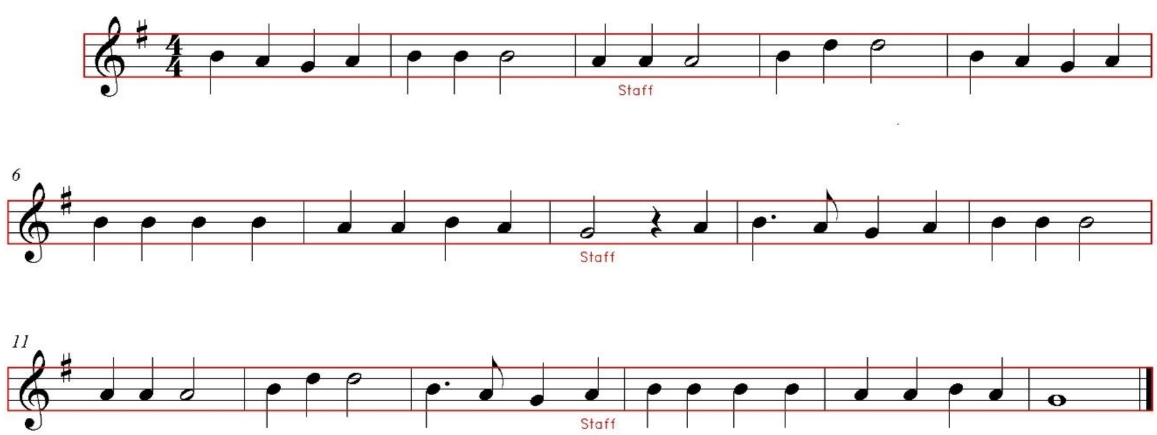
Figure 4.5Projected image

4.3.3Pitch Recognition

The pitch is calculated based on the note head and the distance between the staff lines. First of all, using the Hough transform we detect the position of the staff. By using morphological closing operations hollow notes are filled with solids , and using the template matching method all the note heads are found.



(a)



(b)

Figure 4.6 The staff line and the note heads position

4.3.4Recognition Duration

Neural networks have solved many problems successfully that are difficult to solve in the field of intelligent robots, pattern recognition, predictive inference, etc. Two common neural network methods are used in this paper for symbol recognition. CNNs are very popular applications models for machine vision.

Long Short Term Memory (LSTM) networks are a class of Recurrent Neural Networks (RNNs). Long Short Term Memory (LSTM) networks are widely used for machine learning, including machine translation, tasks involving sequences, speech recognition and text generation. An LSTM Consists of cell, each cells containing multiple gating units that control the removal and addition of information from the state[18].Let x_t, c_t, and $h_t$denote the input, cell, and hidden states, respectively, at iteration t. Given the previous cell state c_t -1, current input $x_t$and previous hidden state h_t-1,the new hidden state ht and the new cell state $c_t$are computed as

$$[f, i, j]^T = [\sigma, \sigma, \sigma, tanh]^T \ ((\mathrm{W}x x_t + \mathrm{U} h_{t-1}) + \mathrm{b})$$

$$c_t = f \ \beta \ c_{t-1} + \mathrm{I} \ \beta \ \mathrm{j}$$

$$h_t = 0 \ \beta \ tanh(c_t)$$

Where b is the bias and showβelement-wise multiplication. The activation function σ is the sigmoid function σ(x) =1/(1 + exp (íx)). The output of an LSTM layer at iteration t is h_t [19].

4.3.5 MIDI Transformation

To record sound MIDI format is a digital music format. MIDI is a technical standard that describes a digital interface, communications protocol and electrical connectors that connect a wide variety of electronic musical instruments, computers and related audio devices for playing, recording and editing music. In this step, for generate the MIDI file we use the MIDIUtil library1. MIDIUtil is a pure Python library within Python programs that allows to write multi-track Musical Instrument Digital Interface (MIDI) files from. The pitch sequence of the previously identified notes is taken as input and duration, and a MIDI file is output as a result. The file can be viewed, edited and music player can be played this file.

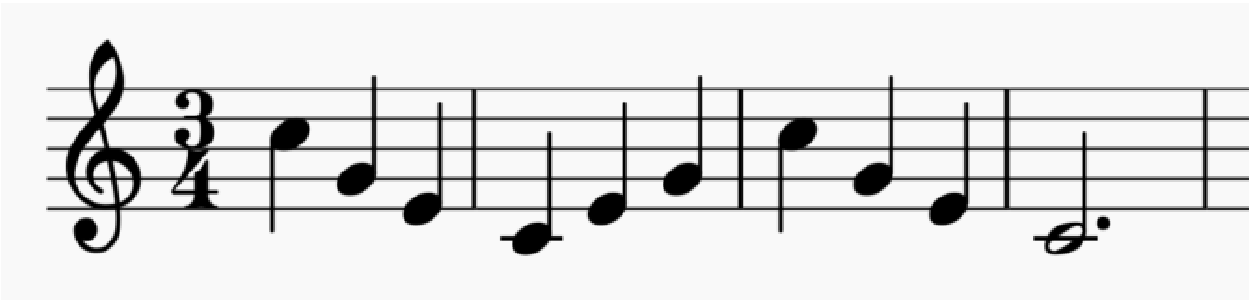


Figure 4.7 MIDI file generated

4.3.6 Optical Music Recognition

We now turn to the problem of optical music recognition (OMR), after presenting the key elements of musical notation, which can be translated into a set of different methods and machine-readable formats for the identification of musical symbols. And translate them into a machine-readable format. Below is an initial strategy within the character recognition framework, which will be discussed in pattern recognition. Before I give an overview of common OMR techniques and applications, I am going to point out the similarities and differences between the fields of OMR and optical character recognition (OCR).

4.3.7 Pattern Recognition

Pattern recognition attempts to classify, assign, and analyze labels on input data. Relatedly, the process is divided into four main stages from the individual approach that goes directly into the capturing phase in our application. Pre-processing was taken to limit the input data to a minimum, which often surrounds the normalization of the input. In the following chapter we will refer to image preprocessing. When performing pattern recognition in images, the components that provide the basis for higher level recognition processes are separated by categories, the properties of the related signals are extractives properties that are classified.

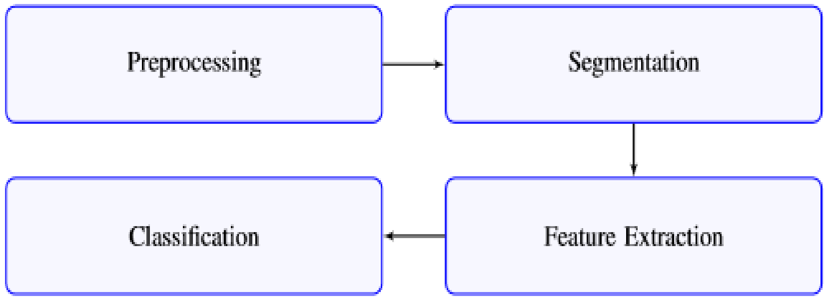


Figure 4.8 Overview of a pattern recognition general system

4.4 Summary

Most music systems make it possible to import and export the final presentation of the musical score for MIDI. However, over the years encoding formats different more music has been created for the music. To overcome this limitation, multiple OMR systems have been studied by merging [13, 55]. However, it is now a challenge. Furthermore only recognition of the highest OMR system results for printed music scores. That's the key difference for state-of-the-art frameworks. This often fails when the input image is in the form of poorly documented photocopies or documents, only during photo scores that work with exception handwriting scores.

Chapter 5

Summary, Conclusion and Implication for Future Study

5.1 Summary of the Study

We will learn from this thesis how to convert the printed music score into MIDI files using the OMR system and what method is used to complete the entire process.

Morphological operations are used to remove the staff lines firstly and for segment the symbols projection method is used . Then neural networks used for recognized the music notes.

5.2 Conclusions

This proposed system has made an effort to representrecognizing a composer's musical signature from bars of music using computer vision & deep learning.

In this thesis, OMR is recommended for printed music scores using CNN for the system. Music score image is first processed to remove staff lines with morphological activity. To segment the note symbols by used projection method and to get the note duration, classified and recognized by their neural networks. The pitch of the notes is recognized according to the distance between the staff line and the note heads. The pitch sequence and duration are converted to MIDI files by MIDIUtil. Finally, the experimental results show that our method yielded excellent results.

5.3 Implication for Further Study

In the future, the proposed method will be used for written score recognition to validate generality, and further notes will be added to the database to test the accuracy of recognition.

References

[1] C Wen, A Rebelo, J Zhang, J Cardoso, “Classi¿cation of Optical Music Symbols based on Combined Neural Network,” International Conference on Mechatronics & Control, 16-10-2019.

[2] A Baró, P Riba, J Calvo-Zaragoza, A Fornés, “Optical Music Recognition by Recurrent Neural Networks,” Iapr International Conference on Document Analysis & Recognition , 15-9-2-2019.

[3] H Miyao, Y Nakano, “Note symbol extraction for printed piano scores using neural networks,” IEICE Transactions on Information and Systems, 20-9-2019

[4] SCJ Dos, A Capela, A Rebelo, C Guedes, dCJ Pinto, “Staff detection with stable paths,” IEEE Transactions on Pattern Analysis & Machine Intelligence,23-9-2019.

[5] HN Bui, IS Na, GS Lee, HJ Yang, SH Kim, “Boosted Stable Path for Staff-line Detection Using Order Statistic Downscaling and Coarse-to-Fine Technique,” International Conference on Pattern Recognition, 28-9-2019

[6] G Chen, S Xia, “The study and prototype system of printed music recognition,” International Conference on Neural Networks & Signal Processing, 2003, 2:1002-1008 Vol.2.

[7] AA Mehta, MS Bhatt, “Optical Music Notes Recognition for Printed Piano Music Score Sheet,” International Conference on Computer Communication & Informatics, 5-10-2019.

[8] QN Vo, T Nguyen, SH Kim, HJ Yang, GS Lee, “Distorted Music Score Recognition without Staffline Removal,” International Conference on Pattern Recognition,8-10-2019

[9] A Rebelo, I Fujinaga, F Paszkiewicz, ARS Marcal, C Guedes, JS Cardoso, “Optical music recognition: state-of-the-art and open issues,” International Journal of Multimedia Information Retrieval, 11-10-2019.

[10] S Dai, Y Tian, CW Lee, “Optical music recognition and playback on mobile devices”, Stanford, 15-10-2019.

[11] DM Dhanalakshmy, HP Menon, V Vinaya, “Musical Notes to MIDI Conversion,” International Conference on Advances in Computing , 2017 :799-804

[12] C Wen, A Rebelo, J Zhang, J Cardoso, “Classi¿cation of Optical Music Symbols based on Combined Neural Network,” International Conference on Mechatronics & Control, 16-10-2019.

[13] AR Blanes, AF Bisquerra, “Camera-based Optical Music Recognition using a Convolutional Neural Network,” Iapr International Conference on Document Analysis & Recognition , 17-10-2019.

[14] M Akbari, AT Targhi, MM Dehshibi, “TeMu-App: Music Characters Recognition Using HOG and SVM,” Machine Vision & Image Processing, 20-10-2019.

[15] J Calvozaragoza, AJ Gallego, A Pertusa, “Recognition of Handwritten Music Symbols with Convolutional Neural Codes,” IAPR International Conference on Document Analysis and Recognition, 23-10-2019.

[16] Y Sazaki, R Ayuni, S Kom, “Musical Note Recognition Using Minimum Spanning Tree Algorithm,” International Conference on Telecommunication System Services and Applications , 28-10-2019.

[17] A Rebelo, G Capela, JS Cardoso, “Optical recognition of music symbols: a comparative study,” Springer-Verlag, 5-11-2019.

[18] H Salehinejad, S Sankar, J Barfett, E Colak, S Valaee, "Recent Advances in Recurrent Neural Networks," arXiv preprint arXiv, 9-11-2019.

[19] S Sen, A Raghunathan, "Approximate Computing for Long Short Term Memory (LSTM) Neural Networks," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 15-11-2019.

[20] W. Zaremba, I. Sutskever, and O. Vinyals. Recurrent neural network regularization. arXiv preprint arXiv:1409.2329, 21-11-2019.

Ashim Report

ORIGINALITY REPORT

4%	0%	4%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1 Ana Rebelo. "Robust optical recognition of handwritten musical scores based on domain knowledge", Repositório Aberto da Universidade do Porto, 2014. — 3%
Publication

2 Ana Rebelo, Ichiro Fujinaga, Filipe Paszkiewicz, Andre R. S. Marcal, Carlos Guedes, Jaime S. Cardoso. "Optical music recognition: state-of-the-art and open issues", International Journal of Multimedia Information Retrieval, 2012 — <1%
Publication

3 Alicia Fornés, Anjan Dutta, Albert Gordo, Josep Lladós. "Chapter 17 The 2012 Music Scores Competitions: Staff Removal and Writer Identification", Springer Science and Business Media LLC, 2013 — <1%
Publication

4 A. Rebelo. "Optical recognition of music symbols : A comparative study", International Journal of Document Analysis and Recognition (IJDAR), 11/17/2009 — <1%

Publication

5	Submitted to University of Bristol Student Paper	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		