Optical Music Recognition: the Case Study of Pattern Recognition

Wladyslaw Homenda

Faculty of Mathematics and Information Sci, Warsaw University of Technology, pl. Politechniki 1, 00-661 Warsaw, Poland,
homenda@mini.pw.edu.pl

Summary. The paper presents a pattern recognition study aimed on music notation recognition. The study is focused on practical aspect of optical music recognition; it presents a variety of methods applied in optical music recognition technology. The following logically separated stages of music notation recognition are distinguished: acquiring music notation structure, recognizing symbols of music notation, analyzing contextual information. The directions for OMR package development are drawn.

1 Introduction

Last two decades are witnesses of brisk development of computing technologies aimed on information and knowledge processing. A special interest of computing technologies development is focused on automation of paper-to-computer memory information transfer. Some paper-to-computer technologies have been successfully developed as, for instance, OCR technology. However, such areas as recognition of printed music, of handwritten text and handwritten music, of geographical maps are examples of fields opened for research and technology development.

In this paper we outline music notation recognition as a case of pattern recognition. The discussion is focused on the three most important aspects of OMR: document structure identification, music symbols' recognition and context knowledge mining. In Section 2 score structure identification is discussed: locating staves, systems, measures and blocks of texts. In Section 3 topics related to music symbols' recognition are discussed. Section 4 includes remarks about music knowledge processing and context information mining.

2 Acquiring Music Notation Structure

Music notation is a highly structured music knowledge container. The structure could be interpreted from two perspectives: logical and geometrical. Music

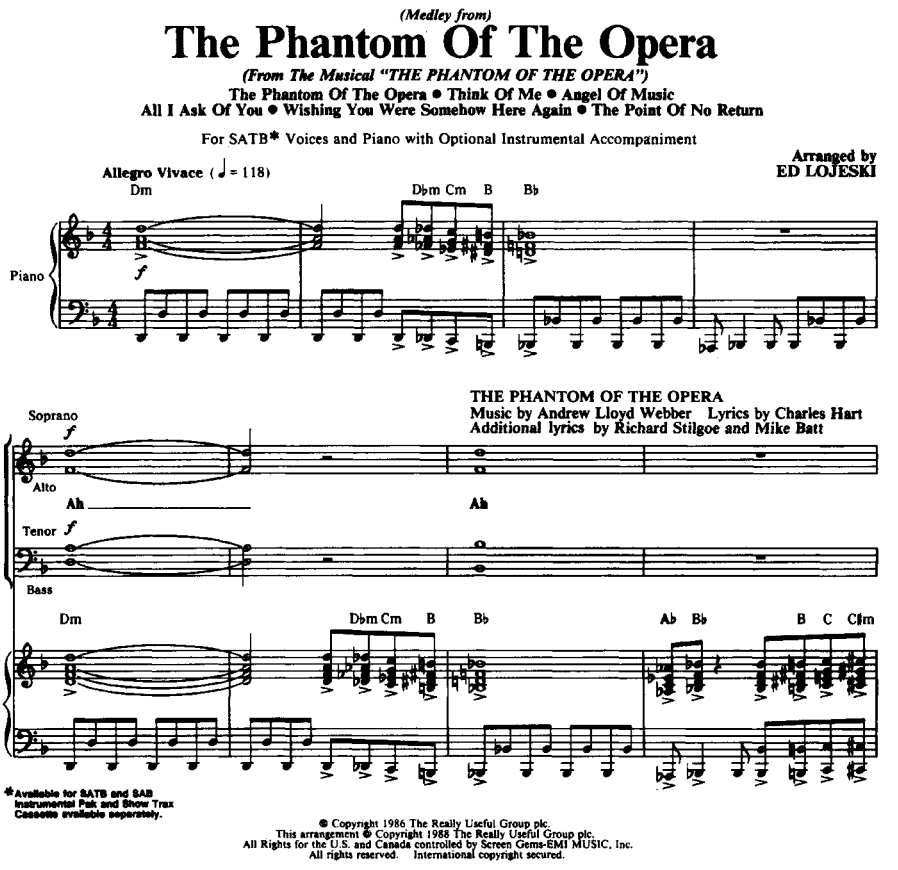

Fig. 1. Score - an example excerpt

notation describes a sequence of events along time axis, usually long sequence of events in long time interval. Conversely, music notation is printed on sheets of paper, so time axis has to be broken in order to fit page width. These two aspects form two perspectives of music knowledge structuring: logical and geometrical structures, c.f. [6] for detailed discussion.

2.1 Logical structuring

Parts and Staves Music notation describes a series of events that could be interpreted as notes and their attributes. These events are placed in three dimensional space. Time and frequency are the obvious parameters forming two dimensions: every note is played in a given time interval and has its own pitch. Music notation usually describes music played by an ensemble. In such a case notes and their attributes related to one instrument create

a part of music and usually fill in one stave, sometimes two or three staves (piano, organ). Separation of music events with regard to instruments creates discrete points of a third dimension of music notation. In this interpretation the time axis is placed along staves while frequency/pitch axis is defined by note position on staff lines and by respective clef. Staves representing parts of music form points of the third dimension. The time axis is scaled in music units: beats and measures where measures are separated by barlines. The set of long staves, not broken by sheets size, form a logical system. Due to its nature logical system consists of staves representing all instruments.

2.2 Geometrical structuring

A piece of music is partitioned into rhythmic entities - measures. In music notation measures are separated by barlines. Besides measure separation barlines play additional roles denoting repetitions of some parts of music and grouping relative instruments as strings or winds. Repetitions are defined by usage of special barlines. Groups of parts are specified by sections of barlines connecting respective staves while groups separation is indicated by broken barlines between groups of staves, c.f. Figure 1.

Pages and systems Music notation, as printed on sheets of paper, is split into parts fitting page's width. The logical system, as described in Section 2.1, is split into parts defined by time intervals which - after necessary stretching - fit page width. This way we get sections of logical system creating systems in the meaning of geometrical structuring. In most cases geometrical systems do not break measures, though it may happen that a long measure is divided between two systems. Unlike logical system, geometrical systems often drop empty staves creating irregular systems, c.f. Figure 1.

Texts Music notation - besides music symbols - also includes texts and alphanumeric characters. The top part of the first page of music score is usually occupied by a kind of music specification: title of the piece, composer and poet names, arranger name, dates, etc. The bottom part of the first page is usually taken by copyright and publisher data. All this information is useless from musical point of view, but are useful for other purposes as, for instance, creating an index of compositions and composers. In contrast, music score often includes lyric as well as separated words or alphanumeric symbols which are important from musical perspective. As examples of such symbols we can take guitar chords, tempo and dynamic markings, fingering, instrument and part names, etc. Locating texts in music notation will give assured advances in further steps of alphanumeric symbols recognition.

2.3 Acquiring music notation structure

Staff lines, long horizontal equidistant lines, are the most characteristic elements of music notation. This feature makes that projections are the most

popular method used for staff lines location, c.f. [3]. Horizontal projections computed on page width will distinguish staff lines with peaks of lines length. Localization of staff lines is easy for ideal notation. However, staff lines of music notation are usually distorted, so then more complex analysis of narrow, local horizontal projections must be done, c.f. [5, 6].

Finding systems is the next step of notation's structure analysis. Since staves of any geometrical system are joined with the beginning barline, finding this vertical line drawn at the beginning of system's staves results in locating system. Vertical projections are usual method applied in localization of barlines what leads to systems location. And again, for distorted notations analysis of local projections avoids all kinds of deformations.

Big systems often have their own internal structure. Staves are clustered collecting instruments of the same group. Such groups may be distinguished with brackets and braces placed at the front of the systems. Therefore, finding brackets and braces is an extension of system location task. Since internal system structuring is also defined by barlines, then both methods can cooperate in inner system's structure location.

3 Recognizing Symbols of Music Notation

Practical recognition process has two crucial stages: image segmentation and symbol classification. The first stage is - roughly speaking - aimed on two tasks: pattern localization and separation patterns from their environment. Localization of patterns most often results in finding bounding box of a given pattern, which is then passed to next steps of recognition process. Recognized patterns may touch each to other, may overlap each other or their bounding boxes may not be disjoint. Separation of such objects significantly simplifies classification and increases recognition rate, c.f. [10]. But objects' separation may be too expensive comparing with recognition rate improvement.

3.1 Segmentation

Since music notation is built around staves, staves' localization indicates placement of other symbols of music notation. Several categories of symbols can be distinguished from the perspective of their placement.

The first class includes symbols that are placed in a strictly determined position. Clefs are always placed at the beginning of stave in fixed vertical position. Key and time signatures appear just at the right sight of clefs while changes of signatures just follow barlines. Key signature always precede time signature if both appear in a measure. Key signature is a sequence of sharps or flats occupying strictly defined vertical position. As a result, finding location of symbols of this class should utilize rules of their placement in addition to image analysis. This kind of domain knowledge facilitates not only localization, but also classification of potential symbols emerging in an investigated

area. Local projections and histograms as well as analysis of vertical and horizontal sections of black pixels are the basic methods utilized in such tasks, c.f. [3, 6]

The second class includes symbols with features that are easy to be detected from analyzed image area. All but whole notes have stems - vertical stick - that could be comparably easy filtered from local vertical projection. Similarly, regions that potentially include sharps, flats and naturals could also be found by vertical projection filtering. On the other hand, all sorts of symbols that may have horizontal sections as arcs (ties and slurs), dynamic hairpins, etc. could be located or partially located by utilizing horizontal projections. Horizontal projections are always affected by staff lines. However, having staves localized, it could be easy remove traces of staff lines from projections.

The third class includes symbols with location determined by other symbols. Among them are so called connectors (beams joining stems of beamed group, symbols of rhythmic groupings, arpeggios), but also accidentals, staccato and other articulation markings, etc. Such symbols usually have their own characteristic features, but also could be placed only according to other, previously found, symbols of music notation. And as above, local projections and histograms as well as vertical and analysis of horizontal sections of black pixels are the basic methods utilized in such tasks.

And, finally, there are symbols that - from recognition perspective, but not music perspective - are randomly placed. Such symbols as rests, change of clefs, dynamic markings, articulation and ornamentation symbols, etc. are examples of irregularly placed symbols. Finding placement of such symbols is often simplified by cleaning the image from already recognized symbols.

3.2 Classification

Compact music symbols Such symbols as clefs, notes, rests, accidentals, signatures, change of clefs have size comparable with stave height. Localization of pattern most often results in finding bounding box of a given pattern, which is then presented to classification module of recognition process. Compact music symbols can be either placed on staff lines or out of them. Therefore, classifiers must deal with random influence of staff lines. Symbols' classifiers cope with a part of original bitmap - bounding box of investigated symbol. Classification decision is made on the basis of direct investigation of bitmap or analysis of extracted features of classified symbol. So, withdrawing from staff lines influence is done either by removing staff lines from original image, by removing staff lines traces from extracted features or by selecting features insensitive to staff lines. Removing staff lines from original image is always run time consuming and may damage other symbols of the image. Since placement of staff lines is usually known, the last two methods overcome the former one.

There is a wide spectrum of features that can be used to characterize classified symbols. These features are extracted from original bitmap restricted by

bounding box of analyzed symbol. Besides the simplest features, as bounding box width/height proportion, more complex features based on projections and moments are utilized, c.f. [7, 11] for detailed discussion on this topic.

As mentioned above, classification can be based directly on original bitmap or on a set of extracted features. Unfortunately, there is no universal classification method that could be successfully applied in music symbols classification. A variety of classifiers are utilized: neural networks, statistical classifiers, centroids and clustering, classification trees, c.f. [7, 8, 10].

Connectors Music symbols may create logical or musical groups. In many cases symbols of such groups are joined with so called connectors. For instance, noteheads of consecutive notes may be connected with ties creating one note of summed total duration, stems of a sequence of eight or shorter notes - with beams instead of having flags. Triplets and other rhythmic groupings outlined as horizontal bracket or arc (with a digit describing rhythmic grouping) are another examples of connectors. Such symbols are usually recognized by investigation of geometrical features of symbols that can potentially be connected. For instance, detection of beams is typically done by exploration of stems' endings opposite to noteheads. Detection of ties is based on finding horizontal arcs, usually flat in their middle parts, which connect consecutive noteheads of the same pitch. Investigation is done by utilization of simple methods: analyzing projections, finding horizontal or slightly sloped line sections, checking geometrical relations between connector and connected symbols.

Non-local symbols Such symbols like slurs, dynamic hairpins and octave modifiers are horizontal or sloped, solid or dashed lines and arcs that cannot be investigated with typical pattern recognizers. Their recognition is based on finding their placement and is usually done as late stage of recognition process. And again, as in case of connectors, non-local symbols are recognized by analysis of such features as local projections, runs of horizontal pixels, tracing their shape, etc. Having score structure identified and other symbols recognized and having position and shape's features of non-local symbols, an analysis of geometrical relations between all of them allows for classification of non-local symbols and - in further processing of acquired music information - describing musical function of them.

OCR Since symbols of music notation vary in size and shapes, some of them can be mistaken with letters. Finding text areas will significantly reduce misrecognitions between alphanumeric and music symbols. Lyric is a distinctive type of text in music notation. It is usually placed under respective staff line. Lyric words are split into syllables with hyphens and underscores extended between syllables. Lyric can appear in one row as well as in several rows. The number of rows may change between systems on a page as well as inside a system. As a result, text location module must cope with such irregularities in order to support recognition of alphanumeric characters.

Recognition of texts obviously applies an OCR technology. Subsequently, adaptation of an OCR package is the simplest solution of texts recognition in music notation. However, such a solution may not be acceptable due to cost of OCR purchasing. As a result, music recognition developer should consider design and implementation of his own OCR package having in mind that domain knowledge would be an important advantage in characters recognition. In [9] a simple technology aimed on text recognition in musical scores is presented. It is based on hierarchical classification method and seems to be adequate for music notation. Some types of music texts use special fonts what makes that it is necessary to apply specialized text recognizer instead of general OCR, c.f. [4].

4 Analysis of Context Information

Analysis of context information is a kind of syntactical pattern recognition. Music notation, despite that is very flexible, must satisfy some strict rules. Of course, due to its flexibility, music notation cannot be restricted by any global description as - for instance - a context free grammar. Nevertheless, there are local rules that could have simple description and are easily verified. Some of such rules were already presented in Section 3. Examples of more complex and very important rules are signatures and voices.

Signatures Key and time signatures put restrictions on the whole piece of music. Time signature defines rhythmic value of consecutive measures (with exceptions of possible upbeat and downbeat). Time signature constraint is strict and does not allow for exceptions. Consequently, music notation recognition - if correct - must satisfy this context constrain. Therefore, time signatures constraint is a tool of recognition's verification and is a possible correction tool. Similarly, key signature affects respective notes in all octaves and its control may be canceled by natural in the octave to the end of the measure of the natural. And as in time signature case, key signature constraint is a verification and - possibly - correction tool of notation recognition.

Voices A part is split to measures along time axis. Alternatively, a part may be divided to voice lines which play important role in music as, for instance, piano extract of orchestral music. The task of voice lines extraction is knowledge processing based on methods of syntactical pattern recognition rather then on optical pattern recognition.

5 Conclusions

Optical music recognition has been intensively developed for last two decades gaining promising results. However, practical realizations in this field are still far from perfection. The field of music notation recognition is still open for

research and further improvements of OMR technology are still sought. This paper gives brief overview of OMR technology from the perspective of pattern recognition paradigm. Three important aspects of recognition process are distinguished: structure of music notation analysis music symbol recognition and context knowledge acquisition. The brief survey of optical music recognition methods is extended for a list of most suitable papers on this subject.

Acknowledgments This work is supported under State Committee for Scientific Research Grant no 3T11C00926, years 2004-2007.

References

1. Bainbridge D, and Bell T (2001) The challenge of optical music recognition, Computers and the Humanities 35:95-121
2. Barton L, W. G (2002) The NEUMES project: Digital Transcription of Medieval Chant Manuscriptis, In: Second International Conference on WEB Delivering of Music, Darmstadt, Germany, IEEE Computer Society Press
3. Fujinaga I (1988) Optical music recognition using projections, MSc thesis, McGill University, Montreal, Canada
4. Gezerlis V, Theodoridis S (2002) Optical character recognition of the Orthodox Hellenic Byzantine Music notation, Pattern Recog. 35: 895-914
5. Homenda W (1996) Automatic recognition of printed music and its conversion into playable music data, Control and Cybernetics, 25(2):353-367
6. Homenda W (2002) Granular Computing as an Abstraction of Data Aggregation - a View on Optical Music Recognition, Archives of Control Sciences, 12(4):433-455.
7. Homenda W, Luckner M (2004) Automatic Recogniton of Music Notation Using Neural Networks, In: Inter. Conf. On AI and Systems IEEE AIS'04, Divnomorskoye, Russia, Sept. 3-10, Proc. Physmathlit, Moscow:74-80.
8. Homenda W, Luckner M (2004) Automatic Recogniton of Music Notation Using Methods of Centroids and Classification Trees, Proc. of the Intern. Symposium ISCIIA'2004, Haikou, China, December 20-24.
9. Homenda W, Luckner M (2005) Hierarchical OCR System for Texts in Musical Scores, submitted, IFSA'2005 World Congress, Beijing, China, July 28-31.
10. Homenda W, Mossakowski K (2004) Music Symbol Recognition Neural Networks vs. Statistical Methods, In: Baets B. et al. (Eds.), Current Issues in Data and Knowledge Engineering, EXIT, Warszawa.
11. Luckner M (2003) Automatic Identification of Selected Symbols of Music Notatio, MSc thesis, Warsaw University of Technology, Warsaw, Poland.
12. Luth N (2002) Automatic identification of music notations, In: Second International Conference on WEB Delivering of Music, Darmsradt, Germany, IEEE Computer Society Press.
13. McPherson J R (2002) Introducing feedback into an optical music recognition system, In: Third Internat. Conf. on Music Information Retrieval, Paris, France.
14. Pinto J C, P. Vieira et al. (2003) A new graph-like classification method applied to ancient handwritten musical symbols, In: International Journal of Document Analysis and Recognition 6(1): 10-22.
15. Rossant F (2002) A global method for music symbol recognition in typeset music sheets, Pattern Recognition Letters 23:1129-41.