Printed Music Recognition

R. Randriamahefa*, J.P. Cocquerez**, C. Fluhr***, F. Pépin**, S. Philipp**

* Institut National des Jeunes Aveugles (Paris France)

Abstract

In this paper, we describe the different steps to recognize printed music. The first step is to detect and to eliminate the staff lines. A robust method based on finding regions where are only the staff lines, linking between them the staff lines pieces in these regions is used. After staff lines elimination, symbols are isolated and a representation called attributed graph is constructed for each symbol. Thinning, polygonalisation, spurious segments cleaning and segments fusion are performed. A first classification, separating all notes with black heads to others, is performed. To recognize notes with black heads (beamed group or quarter notes), a straightforward structural approach using this representation is sufficient and efficient in most cases. In the ambiguous cases (chord or black head linked to two stems), an ellips matching method is used. To recognize half notes and bar lines, a structural method using the graph is used.

1: Introduction

Several works have been performed over the last twenty five years in automatic score recognition. These works have resulted in development of optical printed score recognition [1, 2, 3, 4, 5, 6]. Until now, even the must developped recognition systems can't identify all music notations. These systems have been only tested on some particulars scores. For this reason, we design a large system having a global approach for recognizing scores with simple or multiple voices on one staff including music with chords, with touching symbols or with thin staff lines.

The aim of our work is to develop a system which translate automatically a printed music to Braille music.

Generally, all recognition system is composed of a preliminary stage of segmentation, a stage of attributes extraction and a stage of recognition.

In this paper, firstly we explain the segmentation problems : detection and elimination of staff lines, secondly we discuss the construction of attributed graph representation, thirstly we state the recognition of notes with black heads and at the end we describe the recognition of half note, bar line and other symbols.

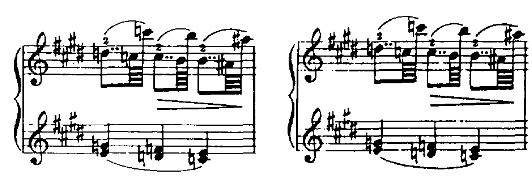
2: Segmentation

The segmentation consists in isolating the symbols in order to identify them. A score is composed of staves and each staff is made up of five lines. The different symbols are writting on or between these lines. Therefore, detecting and eliminating lines are necessary in order to isolate the symbols.

Input image is provided by a Microtek MS300A scanner at resolution 300 dpi. This resolution allows us to obtain an image size 1.1 Mbyte per score page (A4).

2.1: Detection of staff lines

The must straightforward method to detect these lines is the horizontal projection. These lines appear as the projection maxima. This technique suppose that these lines are perfectly horizontal, unbroken and parallel, whereas they are skewed, split and present a slight curvature. These defects appear either at the image acquisition (slope of the score sheet) or at the score edition. Thin horizontal lines of perfect image disappear at the image acquisition (Fig-1a, 1b).



(a) Original score (b) Scanned score
Fig.1 An example of score with thin line

Therefore, another technique taking into account these defects may be designed. The Hough transform using on little part of image (to limit processing time) give a satisfactory result for perfect scores but a bad result is obtained for slanted scores.

The method we have retained is composed of several stages:

0-8186-4960-7/93 $3.00 © 1993 IEEE

- vertical projection,
- projection filtering,
- local minima regions finding,
- horizontal projection of each local minima regions,
- linking different peaks beteween them in order to build up the staff lines.

From the filtered vertical projection, local minima regions are finding. These regions correspond probably to the regions where there is only the staff lines. Then, they are validated according to their height and their width.

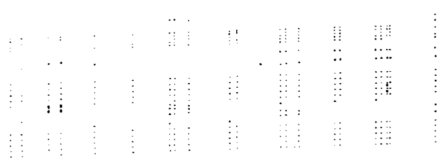
Once these regions have been found, a little part of each region is considered for the horizontal projection so that the peaks of this projection correspond certainly to these lines. The peaks which their width are greater than a threshold are eliminated. This threshold is the average of peaks widths in the region. These spurious peaks are rejected in order to have only staff lines pieces, hence accurate equations of staff lines will be obtained. These large peaks can be beams pieces.

Having found the horizontal projection peaks of all regions (Fig-2b), these must be combined in order to determine staff lines. Two solutions have been studied:

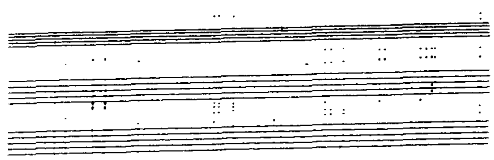
- the first, using the Hough transform method, has been tested. Good results have been obtained for some scores, but for bowed or fragmented lines the result is not satisfactory.
- the second, which was retained, utilizes the fit line square method to combine the peaks belonging to the same line. The threshold, used by this method, is not severe so that the peaks belonging to the bowed line are linked.

The possible lines have been found, and two validations allow us to keep only the staff lines (Fig-2c). This technique is robust because even if these lines are bowed, skewed or fragmented they have always been found.

(a) Score



(b) Horizontal projection peaks



(c) Staff lines detected

Fig.2 Detection of staff lines

2.2: Elimination of staff lines

Once staff lines have been found (lines equations determined), pixels of lines can be eliminated so that the different symbols are isolated. The elimination of staff lines, based on thickness criterion, poses a problem. A few symbols are splitted since they are tangent to these lines. Differents works have tried to solve this problem (for instance with a neural network [2]).

We have used the straightforward criterion based on thickness. The thickness of each line is estimated in according to the width values of the lines peaks. Knowing the staff line equation, the staff line width at each point is determined. If this width is smaller than a threshold (proportionnal to the estimated line thickness), the line points at this place are erased. The problem of split symbols is carried over in recognition stage.

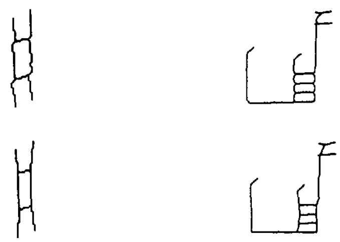
3: Construction of attributed graph representation

Once staff lines have been eliminated, symbols are isolated by using growing regions method and thinning by the improved Zhang and Suen algorithm [6] is performed in order to facilitate subsequent processing and to apply structural method for recognition. This skeletonization algorithm has been proven to be insensible to slight variations of contour.

After thinning, the construction of attributed graph is performed.

3.1: Polygonalisation

First, the different particular points of skeleton are marked (end, junction points). A line tracking technique is applied to determine the different skeleton parts. Each skeleton part is defined between two particular points (end or junction point). Next, the parts are approximated by piecewise straight line segments. We have used the method of the chord, with a variable threshold calculated from an exponential function of number of part pixels and the staff space. This function is different for small and large size objects. For a small size object, a skeleton long part having a slight curvature is composed of single part and for a large size object, this slight curvature can be considered, hence this skeleton long part will be split in several straight lines (Fig-3a, 3b). These functions have been determined empirically.



(a) Small object (b) Large object
Fig.3 Skeleton and Polygonalized symbols

3.2: Minimum distance contour attribution

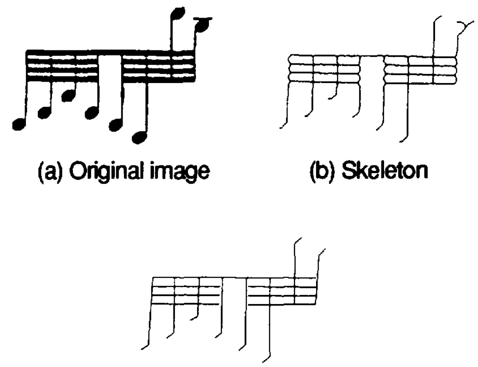
After polygonalisation, a parameter defined as a minimum distance to the contour is attributed to each segment. During thinning, we have attributed to each pixel its minimum distance to the contour (equal to the iteration rank which the pixel is kept). Then, the distance to the contour attributed to each segment is determined from the values attributed to its points. The importance of this parameter is to allow us to distinguish segments belonging to full ellips. The aim is to attribute the maximum distance of its points to the note head segments (full ellips) and the minimum to others. Analyzing the skeleton map distance, we have noticed that the distance values in note head segments (full ellips: black heads) vary quickly, on the other hand these distance values are pratically unchanging for others segments. From this observation, the average of chord slopes, joining the current pixel at the one segment extremity is used to evaluate this variation.

After polygonalisation, spurious segments may be eliminated and some configurations of segments set are transformed. The criteria used by this process depend on both canditate segment and neighbours parameters. These parameters are:

- the type of extremities points (end, junction and normal point),
- the slope,
- the distance to the contour.

This process has been performed in order to simplify the skeleton structure while keeping the essential segments and to make the graph representation to be close to the original morphology of the object.

The attributed graph is constructed : the graph nodes correspond to the segments and the graph arcs to the links between segments (Fig-4).



(a) Original image (b) Skeleton

(c) Result
Fig.4 Graphical representation of the constructed attributed graph

After constructing the graph, pre-classification is performed. By a simple criterion using the distance to the contour, we classify the symbols into two categories: the symbols including notes with black heads and the others.

In the first category, there is at least one segment having a distance to the contour exceeding a threshold. This threshold is proportionnal to the staff space.

4: Recognition of notes with black heads

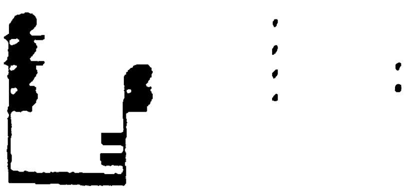
A beamed group note is composed of basic notes (eighth, sixteenth note, etc...). Then, there is an infinite combination. Hence, in order to recognize the beamed group and quarter notes, a simple structural approach consisting in basic notes identification is sufficient. The recognition is thus limited to find the basic note structure. Each basic structure is composed of some primitives. A quarter note is consisting of a head and a stem. A beamed group note is consisting of : heads, stems and hooks. Each primitive has its distinctive features:

- head: its segment slope vary between 0° to 45°, its distance to the contour exceeds the threshold described above, its length is smaller than a threshold proportionnal to the staff space and it is linked to a single stem,
- stem: its segment is vertical (slope between 80° and 100°) and its distance to the contour is very small,
- hook: its segment slope and its distance to the contour are both small, and it is linked to one or several stems.

In some cases, some heads don't check these features. This case appears when several heads are put together (chord) or a head is linked to the two voices. Therefore, the skeleton of two neighbour heads corresponds to a loop or a long segment. Ellips matching is performed in the regions when the types of ambiguity are found. The size of ellips model is calculated from the staff space. So that this matching method is robust, this size is taken to be slightly smaller than these in the score. Its orientation is taken to 30°. Once, heads in ambiguous regions are detected (Fig-5), the stems which heads are linked are founded.

From the graph, the segments identified as stems, hooks or black heads are removed, the remaining sub-graph will be treated afterwards. In the case, the other symbols connected to the beamed group note can be recognized.

The reason, which the ellips matching method is not applied to the global image in order to detect all black heads, is a time consuming nature.



(a) Original image (b) Detected notes
Fig.5 Chord notes

5: Recognition of half notes, bar lines and the other symbol

Structurally, a half note is composed of stem and loop. The half note recognition consists of following steps:

- detection of stem: in the attributed graph, the stem is a vertical segment (slope between 80° and 100°), its distance to the contour is very small, its lenght exceeds a threshold proportionnal to staff space,
- detection of little loops: the height of the loop is close to the staff space.

In the score, the bar line corresponds to a vertical line which its height is equal to the neighbour staves height (printed piano music) or to the staff height. This feature allows us to recognize easily the bar line. Two close bar lines are classified as double bars. Repeat marks is a double bars with two dots at second and third space. Sometimes, bar line is connected to another symbol (slur, tie, etc...), the segments part of bar line are removed from the graph, the remaining sub-graph will be considered to be others symbols.

To recognize others symbols, which the shape is invariable, several techniques based on learning are usable. We will use the neural network learnt with the backpropagation method. The network input is constructed from the attributed graph.

6: Conclusion

The staff lines equations determined by the technique described are accurate. This precision is important in order to find the true pitch of notes. We have noticed that the approach described to recognize printed music symbols is hierarchical. This method is suitable for this problem since some symbols are totally different. The recognition rate of notes with black heads is nearly 100 % for some parts of different scores. Our recognition system doesn't use yet a syntactic analysis which will be added to the system in order to increase its performance (triplet idenfication, identification of note belonging to the same voice, correcting the ambiguous recognition). The veritable problem of printed music recognition lies in touching symbols.

* Institut Nationale des Jeunes Aveugles (Paris France)

** Ecole Nationale Supérieure de l'Electronique et de ses Applications (Cergy Pontoise - France)

*** Institut Nationale des Sciences et Techniques Nucléaires (Centre d'Etudes Nucléaires - Saclay France)

References

[1] H.Kato and S.Inokuchi, "A Recognition System for Printed Piano Music Using Musical Knowledge and Constraints," Proceedings of the International Association for Pattern Recognition workshop on Syntactic and Structural Pattern Recognition, Murray Hill, New Jersey USA, June 13-15, 1990.

[2] P.Martin and C.Bellissant, "Low-Level Analysis of Music Drawings Images," Proceedings of the First International Conference on Document Analysis and Recognition, Saint-Malo, France 1991.

[3] T.Matsushima, T.Harada, I.Sonomoto, K.Kanamori, A. Uesugi, Y. Nimura, S.Hashimoto and S.Ohteru, "Automated Recognition System for Musical Score - The Vision System Wabot-2," Bulletin of Science and Engineering Research Laboratory, Waseda University, 112, September 1985, pp. 25-52.

[4] A.T.Clarke, B.M.Brown and M.P.Thorne, "Using a Micro to Automate Data Acquisition in Music Publishing," Microprocessing and Microprogramming 24 (1988) 549-554.

[5] I.Fujinaga, B.Pennycook and B.Alphonce, "Computer Recognition of Musical Notation," The First International Conference on Music Perception and Cognition, Kyoto, Japan, 17-19 October 1989.

[6] C.M.Holt, A.Stewart, M.Clint & R.H.Perrott, "An improved parallel thining algorithm," Communications of the ACM, Vol 30, N° 2, Fev. 1987.