Staff and graphical primitive segmentation in old handwritten music scores

Alicia Fornés [a,1], Josep Lladós [a] and Gemma Sánchez [a]
[a] *Computer Vision Center, Computer Science Department*

Abstract. In document analysis field, Optical Musical Recognition is a mature area in printed scores, whereas few research works have been done in handwritten ones. The difficulties in handwritten scores are increased if we work in old documents, because of paper degradation and the lack of a standard in musical notation. In this paper we propose a method to segment staff and graphical primitives in old handwritten scores. The extraction of staff lines has been performed using Hough Transform, skeletization, median filters and a contour tracking process. The segmentation of lines and head notes has been done using morphological operations and median filters. Our method has been tested with several scores of XIX century with high performance rates.

Keywords. Document Analysis, Graphics Recognition, Optical music recognition, Old documents, Handwritten diagrams

1. Introduction

Document Image Analysis (DIA) is a Computer Vision discipline that combines Pattern Recognition, Image Processing and Artificial Intelligence techniques to the automatic recognition of composing elements and knowledge extraction from document images. Two major research trends are active in DIA, namely the extraction of textual entities or Optical Character Recognition (OCR), and the extraction of graphical ones or Graphics recognition (GR). In this paper we are focused in Optical Music Recognition (OMR), whose aim is the identification of music information from images of scores and their conversion into a machine legible format. This process allows the development of a wide variety of applications: edition and publication of scores never edited, renewal of old scores, conversion of scores into Braille code, creation of collecting databases to perform musicological analysis and finally, production of musical description files: NIFF (Notation Interchange File Format) and MIDI (Music International Device Interface).

Although Optical Music Recognition has close relationship to OCR, it is considered a Graphics Recognition problem because it requires the understanding of two-dimensional relationships. To help in the recognition process, grammar rules can be formalized from context information of musical notation.

There is a lot of literature about recognition of printed scores. A survey of classical OMR (from 1966 to 1990) can be found in [1], where several methods to segment and

[1]Correspondence to: Computer Vision Center, Universitat Autònoma de Barcelona, 08193, Bellaterra, Barcelona, Spain. Tel.: +34 93 581 18 28; Fax: +34 93 581 16 20; E-mail: afornes@cvc.uab.es.

recognize symbols are reviewed. For detecting staff lines, projections, slicing techniques, classifiers (based on decision trees) and line adjacency graphs are usually used, whereas graph grammars and matching methods are commonly used to classify musical symbols.

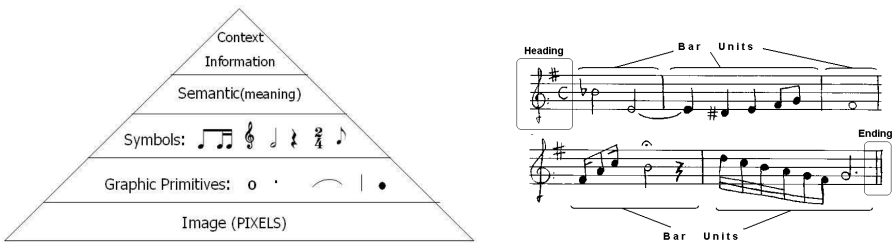
Few research works have been done in handwritten scores [2,3], which ones introduce additional difficulties in the segmentation and the recognition process: notation varies from writer to writer, symbols are written with different sizes, shapes and intensities; the number of touching and broken symbols increases significantly.

Another growing interest in the Document Analysis area is the recognition of ancient manuscripts and their conversion to digital libraries, towards the preservation of cultural heritage. In this work we combine the above-mentioned requirements: the recognition of old handwritten musical scores. We are focused on the recognition of old handwritten scores (XVIII-XX centuries) so that these scores of unknown composers could be edited and published (contributing to the diffusion of artistic and cultural heritage). Old scores make the recognition task more difficult, because of paper degradation (scores in poor condition) and the lack of a standard notation. In addition, there are distortions caused by staff lines, broken and touching symbols as well as high density of symbols. For those reasons, an expert system will be required to learn every new way of writing, and artificial intelligence based techniques will take advantage of higher level musical information.

In this paper we focus on the early stages of the system (extraction of staff and graphical primitives), organized as follows: In section 2 an architecture of the system and the structure of musical scores are presented. Section 3 describes our approach to segment score elements: extraction of staff lines, vertical lines, and head notes. Section 4 reports some illustrative experimental results. Finally, the concluding remarks are exposed.

2. Layers and structure of musical scores

As we can find in [4], an OMR system has five layers, corresponding to the five abstraction levels of the processed information, see Fig. 1.a: the first one is the image one, formed by pixels; then we can find the graphical primitive layer (where this paper is focused on) formed by dots, lines, circles and curves. In the symbol layer graphical primitives are combined to form musical symbols, such as notes and rests. In the semantic-meaning layer information, the pitch and the beat is obtained. Finally, in the context information layer, grammar rules are used to validate and solve ambiguities.



(a) (b)

Figure 1. (a) Levels. (b) Structure of a score.

Feedback among layers is extremely important because each level contains hypothesis of various levels of abstraction, so, if an upper layer reject a result produced from lower layers (e.g. a certain object is not what it has been determined to be), the system must be able to correct this error and classify the object again.

The musical notation consists of the following elements: staves (where musical symbols are written down), attributive symbols at the beginning (clef, time and key signature), bar lines (which separate every bar unit) that include rests and notes (composed of head notes, beams, stems, flags and accidentals); and finally, slurs and dynamic and pedal markings. Some scores include text, so an important task is to determine which objects are text (lyrics), and which are musical symbols. In addition, some words correspond to dynamic markings, so context information should help to distinguish them.

Formal language theory provides useful tools to recognize and solve ambiguities in terms of context-based rules. Grammars are usually used to describe the score structure, see Fig. 1.b). Therefore, parsers guide the recognition and validation process. Informally speaking, a grammar describing a score consists of three blocks **G: S -> H[B]E**, where **H** is the heading with the attribute symbols: treble, alto or bass clef, time signature (commonly formed by two numbers that indicate the measure) and key signature (flats, sharps or naturals, which indicate the tonality of the score). All these symbols are very important to provide the meaning of musical symbols. Then, the score is discomposed in bar units **B**, in which notes and rests are written down. The amount of notes and rests in every bar unit depends on the time signature, so it will obviously help to solve ambiguities in the recognition of notes and rests. Finally, there is an ending measure bar (**E**).

Our recognition strategy follows a typical OMR architecture: After preprocessing the image, a segmentation process extracts graphical primitives; then recognition and classification of musical symbols is performed. Finally, context information is used to validate it and solve ambiguities. The segmentation of staff lines and graphical primitives is described in the following section.

3. Segmentation method

After scanning the image (at a minimum resolution of 300 dpi), it is binarized using Niblack locally adaptive binarization method [5] because global binarization techniques do not work with such degraded old images. Then, the image is deskewed using Hough Transform to detect the orientation of the score.

3.1. Extraction of Staff lines

Because of deviations in staff, the recognition of staff lines is a complex task in old handwritten scores. Staff lines are perfectly horizontal in printed scores, so horizontal projections can be effectively used to detect them [6]. Working with old handwritten scores demands the usage of more sophisticated techniques, because staff lines rarely appear completely straight and they often present gaps in between. This is caused by the degradation of old paper, the warping effect and the inherent distortion of handwritten strokes when the staff lines are drawn by hand.

Our approach consists in obtaining a coarse approximation of staff lines location (with horizontal projections), and performing a contour tracking process: After performing a run length smearing process, horizontal runs are used as seeds to detect a segment of every staff line; then it is tracked in both directions following the best fit path according to a given direction. In order to avoid deviations (wrong paths) in the contour tracking process, it is necessary to consult a rough staff approximation, which has been obtained joining horizontal aligned segments from the skeleton of the image. Using median filters with a horizontal element, most symbols are deleted from the skeleton of the image, and only those symbols with horizontal shapes will remain in the image. The size of this horizontal structuring element is constant, because in the skeleton image, each line is one pixel-width, so the width of lines in the original image is irrelevant. Once we have the image with horizontal segments, each segment is joined with others according to its orientation, distance and area. Figure 2.a shows an original score suffering from a warping effect and its staff reconstruction (fig. 2.b).

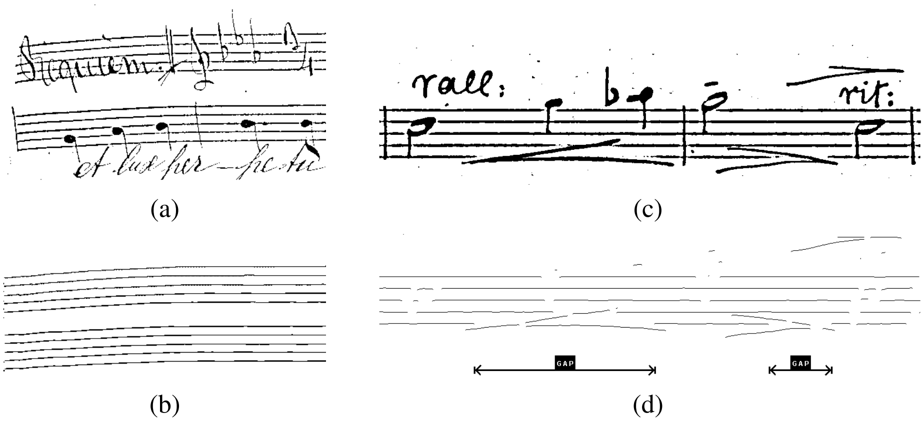


Figure 2. (a) Original Image (b) Reconstruction of staff lines. (c) Original Image (d) Line segments of staff lines with gaps and horizontal symbols

Our prototype has troubles with big gaps (in staff lines) in presence of symbols with horizontal shapes. Figure 2.c shows a big gap with a crescendo marking, so in fig. 2.d we can see that the staff reconstruction process could fail following a segment of the crescendo marking instead of jumping to the next segment of staff line. A solution to this problem consists in increasing the size of the slice, but it could not work in scores with large deviations in staff lines.

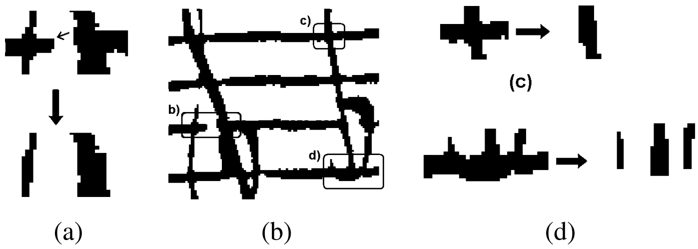


Figure 3. Examples of Line Removal in Contour Tracking process. a) Original Image, b) Gap in line, c) Symbol crosses the staff line, d) Symbol is tangent to staff line. Symbol becomes broken

Once we have obtained the reconstructed staff lines, the contour tracking process can be performed as it has been described before. If there is no presence of staff line (gap), the contour tracking process will be able to continue according to the location of the reconstructed staff line. Concerning line removal, we must decide which line segments can be deleted from the image, because if we delete staff lines in a carelessly way, most symbols will become broken. For that reason, only those segments of lines whose width is under a certain threshold will be removed. This threshold depends on width of staff lines, calculated using the statistical mode of line-segments. Figure 3 shows some examples of line removal: Figure 3.a is the original image, where in Fig. 3.b we can see how in presence of a gap, the process can detect next segment of staff line to continue; in Fig. 3.c a symbol crossing the line will keep unbroken, because the width of the segment is over the threshold.

In this level of recognition, it is almost impossible to avoid the deletion of segments of symbols that overwrite part of a staff line (they are tangent to staff line, see fig. 3.d and whose width is under this threshold, because context information is not still available.

3.2. Recognition of vertical lines

After deleting staff and calculating distance between stave lines, vertical lines and head notes are the first graphical primitives to recognize. First, some morphological operations and run length smearing techniques are used to reduce noise. Afterwards, we use median filters with a vertical structuring element, so only symbols with vertical shape will remain (see Fig. 4). Contrary to extraction of staff lines, here the size of the structuring element depends on the distance between staff lines. We have also tested Hough Transform to detect vertical lines, but results are very similar and the algorithm is slower.

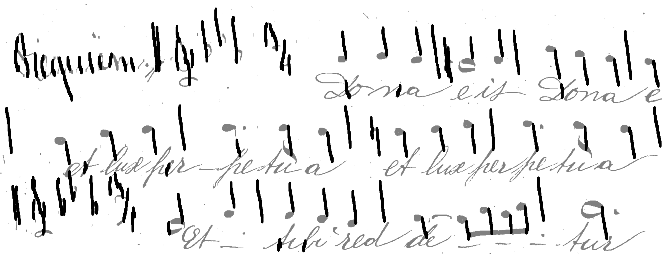


Figure 4. Vertical lines detected are in black color.

3.3. Recognition of filled head notes

Working with printed scores makes this process easier, because all head notes have similar shape. A morphological opening operation (with a circular structuring element), and choosing the ones with adequate circularity and area, does not work with handwritten scores, because there is too much variability in ways of writing to perform a process that detects exactly all head notes. The method proposed performs a morphological opening with elliptical structuring element (whose size depends on the distance between staff lines), oriented 30 grades. After that, elements with large area are discarded. This approach gets all filled head notes and false positives (see fig. 5), but it is better to discard false positives in next stages than forgetting real head notes.

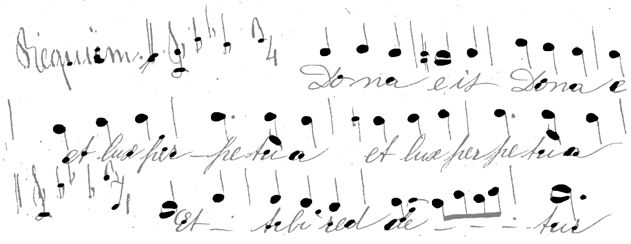


Figure 5. Filled head notes detected in black color.

Because of the lack of a standard notation in old scores, some modern rules of musical notation are not applied in old scores: E.g. in modern musical notation, if a head note is allocated under the third staff line, it has a beam on the right side; otherwise, it has a beam on the left. In some old scores, notes have the beam on the right side whatever their allocation on the staff. So, we will classify notes (filled head notes with beams) in higher-level stages, using grammar rules and the knowledge of time signature.

3.4. Recognition of bar lines

Once we have detected vertical lines and filled head notes, lines must be classified, see Fig. 6, in beams (which have one or more head notes), bar lines (longer than beams, without head notes) and others (e.g. lines that are part of another kind of symbols, such as flats, sharps or naturals). Bar lines are the most important vertical lines, because they divide scores in bar units. Once we have isolated every bar unit, we can process them in an independent way, looking for musical symbols using grammar rules.

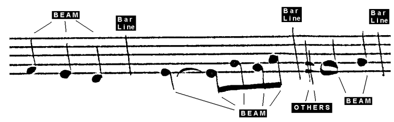


Figure 6. Verticals in scores.

A first approximation of bar lines is performed assuming these two hypotheses: bar lines cover all staff and there are no head notes in their extremes. So, if a vertical line is large enough and it is situated covering all five staff lines, then we will label it as a bar line if there is no presence of filled head notes in its extremes, see Fig. 7.

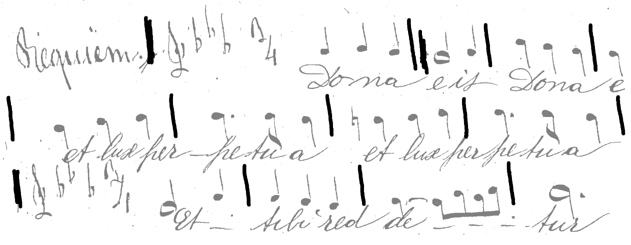


Figure 7. Bar lines in black color.

Scores	Verticals: Correct / Detected, (% FP)	Bar lines / % FP	Head notes / % FP
1	39 / 51 , 30%	5 / 8 , 25%	15 / 42 , 65%
2	51 / 63 , 24%	7 / 9 , 22%	10 / 28 , 64%
3	46 / 64 , 28%	6 / 7 , 15%	23 /68 , 66%
4	42 / 60 , 30%	8 / 8 , 0%	20 /63 , 69%
5	39 / 58 , 32%	7 / 8 , 25%	31 /88 , 64%
6	36 / 49 , 27%	9 / 11 , 18%	12 /40 , 70%
7	20 / 31 , 35%	8 / 10 , 20%	37 /97 , 62%
8	32 / 42 , 23%	8 / 8 , 0%	29 /85 , 66%
9	24 / 37 , 35%	6 / 7 , 15%	15 /47 , 68%

Table 1. Results: 100% of Head notes, Vertical and Bar lines detected. FP= False Positives

4. Results

We have tested our method with a set of images of several composers. These images of scores have been obtained through the archive of Seminar of Barcelona. In last section we have shown some results from a section of the Requiem Mass of the composer Aleix. Figure 8 shows results from a section of "Salve Regina" of the composer Aichinger are shown: Staff is removed, head notes and vertical lines are in black color; bar lines are shown as the thickest vertical lines. The remaining score is in grey color.

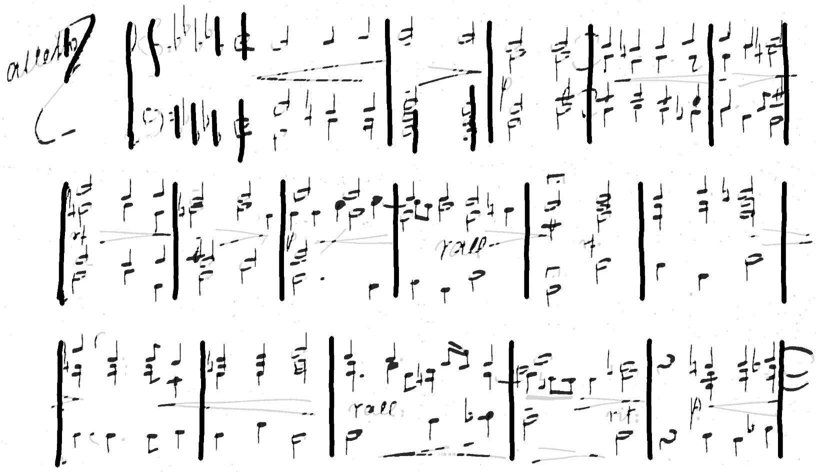


Figure 8. Bar lines in black color.

Several staffs of scores from different composers have been tested. In table. 1 we can see that head notes, vertical and bar lines detected and the percentage of false positives (which will be detected in high-level layers).

5. Conclusions

In this work an approach to segment primitive elements in handwritten old music scores has been presented. Among the tested methods for this early stage of segmentation, the ones with higher performance rates have been chosen. Our strategy consists of the following steps: First, score line detection and removal, using hough transform and a line tracking algorithm. Second, the detection of vertical lines using median filters. Third, circular primitives corresponding to filled head notes have been extracted using morphological operators. Finally, bar lines have been detected from reconstruction of staff lines and head notes.

Our work is in a preliminary stage, but we have obtained high performance rates in this primitive segmentation stage. False positives in the recognition process are due to the enormous variation in handwritten notation and the lack of a standard notation.

Further work will be focused on improving the reconstruction of staff lines and obtaining other graphic primitives and combining them to classify musical symbols: To deal with the reconstruction of staff lines in presence of big gaps in a more robust way, we could create a graph with segments as nodes, and use a shortest path algorithm (with backtracking) to reconstruct it. The detection of whole and half notes is more difficult than filled head notes, because handwritten circles are often broken or incomplete, so morphological operations cause too many false positives. To cope with the extraction of text (lyrics), we can assume that words have more curvature than musical symbols, so structural tensor algorithm could help in their recognition. The recognition of attribute symbols at the beginning of the score (key, clef and compass signature) will require an expert system to learn every way of writing. Finally, a grammar will be formalized to help in the classification of musical symbols.

Acknowledgements

We would like to thank Josep Maria Gregori Cifré from Art Department of UAB for his help in accessing to old resources of archive of Seminar of Barcelona.

This work has been partially supported by the project TIC2003-09291.

References

[1] Blostein D., Baird H., "A Critical Survey of Music Image Analysis," in Structured Document Image Analysis, Eds. H. Baird, H. Bunke, and K. Yamamoto, Springer Verlag, 1992, pp. 405-434

[2] Ng K. C., "Music Manuscript Tracing", in Proceedings of the Fourth IAPR International Workshop on Graphics Recognition (GREC 2001), Kingston, Ontario, Canada, pp. 470-481.

[3] Pinto, J. C., Vieira P., Sosa, J. M., "A new graph-like classification method applied to ancient handwritten musical symbols", International Journal of Document Analysis and Recognition (IJDAR 2003) Vol. 6, Issue 1, pp. 10-22.

[4] Kato, H. and S. Inokuchi (1990). "The recognition system for printed piano music using musical knowledge and constraints". Proceedings of the IAPR Workshop on Syntactic and Structural Pattern Recognition, Murray Hill, New Jersey.

[5] Niblack, W. "An introduction to digital image processing". Englewood Cliffs, Prentice Hall, 1986, pp.115-116

[6] Bainbridge, D. and Carter, N. (1997), "Automatic Reading of Music Notation, in: Handbook of Character Recognition and Document Image Analysis, eds. H.Bunke and P.S.P.Wang, World Scientific, Singapore, pp.583-603.