Primitive segmentation in old handwritten music scores *

Alicia Fornés[1], Josep Lladós[1], and Gemma Sánchez[1]

Computer Vision Center / Computer Science Department,
Edifici O, Campus UAB 08193 Bellaterra (Cerdanyola), Barcelona, Spain
{afornes,josep,gemma}@cvc.uab.es
http://www.cvc.uab.es

Abstract. Optical Music Recognition consists in the identification of music information from images of scores. In this paper, we propose a method for the early stages of the recognition: segmentation of staff lines and graphical primitives in handwritten scores. After introducing our work with modern musical scores (where projections and Hough Transform are effectively used), an approach to deal with ancient handwritten scores is exposed. The recognition of such these old scores is more difficult due to paper degradation and the lack of a standard in musical notation. Our method has been tested with several scores of XIX century with high performance rates.

1 Introduction

The aim of Optical Music Recognition (OMR) is the identification of music information from images of scores and its 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 audio files or musical description files: NIFF (Notation Interchange File Format) and MIDI (Music International Device Interface).

Although OMR has many similarities with Optical Character Recognition (in fact OCR is a sub-task of OMR because lots of scores include text), OMR requires the understanding of two-dimensional relationships. It is nevertheless true that music scores follow strict structural rules that can be formalized by grammar rules, so context information can be extracted helping in the recognition process. A survey of classical OMR (from 1966 to 1990) can be found in [1], where several methods to segment and recognize symbols are reviewed: the detection of staff lines is performed using projections, a line adjacency graph, slicing techniques, comparing line angle and thickness. Extraction and classification of musical symbols is performed using projections, classifiers based on decision

* This work has been partially supported by the Spanish project CICYT TIC 2003-09291

trees, matching methods and contour tracking properties. Finally, validation of scores is usually done using grammars.

OMR is a mature area for printed scores, however our work is focused on the recognition of handwritten ones: we propose a method to detect primitives in modern and old handwritten scores. In modern ones, the detection of staff lines is performed using Hough Transform and projections, whereas in old scores, a contour tracking process is required to cope with deviations in staff. Concerning graphical primitive detection, we propose similar approaches either for modern and old scores: morphological operations, Hough Transform and median filters.

This paper is organized as follows: In section 2 the structure of scores and layers of the system are shown. In section 3 our work with modern handwritten scores is presented, whereas our approach to the segmentation and classification of primitive score elements in old handwritten scores is described in section 4. In section 5 some illustrative experimental results are reported. Finally, in section 6 the concluding remarks are exposed.

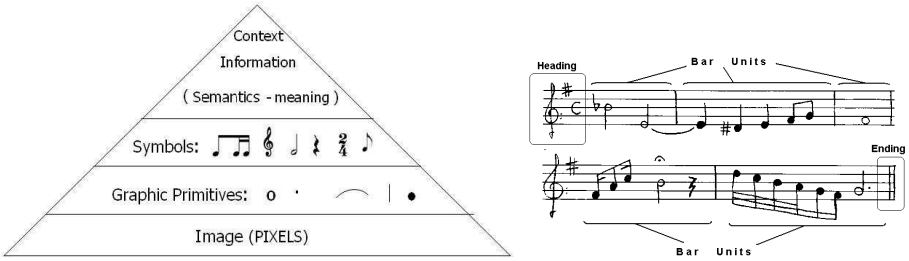
2 Handwritten scores: Structure and Layers

Whereas there is a lot of literature about the recognition of printed scores, few research works have been done in handwritten ones [2, 3]. Regarding printed ones, handwritten scores 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.

According to the approach proposed by Kato [4], an OMR system has several layers, corresponding to the abstraction levels of the processed information, see Fig. 1(a): the image layer is formed by pixels; the graphical primitive layer is formed by dots, lines, circles and curves. In the symbol layer, graphical primitives are combined to form musical symbols. In the semantic-meaning layer information, the pitch and the beat of every note is obtained, and grammar rules are used to validate it and solve ambiguities. Feedback among layers is extremely important because each level contains hypothesis of various levels of abstraction, so, if an upper layer rejects 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 in scores consists of the following elements: staffs (when musical symbols are written down), attributive symbols at the beginning (clef, time and key signature), bar lines (that separate every bar unit) that include rests and notes (composed of head notes, beams, stems, flags and accidentals); and finally, slurs and dynamic 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 or semantic restrictions using attributes.



(a) (b)

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

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** $\rightarrow$ **H[B]E**, where **H** is the heading with the attribute symbols. Then, the score is discomposed in bar units **B**. The end of the score is marked with an ending measure bar (**E**).

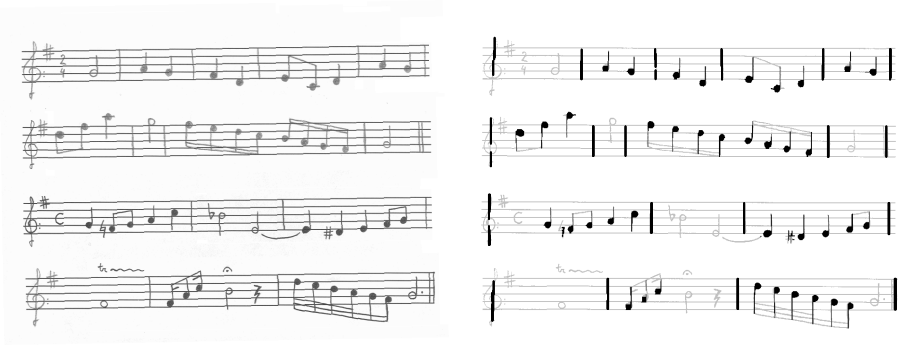
Our recognition strategy follows a typical OMR architecture: After preprocessing the image, a segmentation process extracts graphical primitives; then recognition and classification of musical symbols is performed. Finally, a semantic layer uses context information to validate it and solve ambiguities.

The early-level stages are described in this paper: segmentation of score blocks and detection of primitives. As we have said before, most segmentation problems are due to distortions caused by staff lines, broken and touching symbols as well as high density of symbols. For this reason, deleting staff lines and isolating symbols are the first tasks to cope with.

3 Modern Handwritten scores

Initially, we have been working with modern musical scores, where paper is in good condition, there is a standard of musical notation and most of staff lines are printed. Here, the approach proposed consists in the following: First, the input image (at a resolution of 300 dpi) is binarized(using the OTSU method) and deskewed (using Hough Transform to detect staff lines). After that, horizontal projections can effectively be used to detect rows likely to contain a staff line. In the staff analysis some parameters are set: width of staff lines and distance between them. Knowing these parameters, a run-length smearing process deletes staff lines trying to keep complete symbols. Finally, morphological operations reduce noise.

Concerning the primitive detection stage, vertical lines and head notes are the first graphical primitives to recognize: detection of vertical lines is also performed using the Hough Transform (allowing a skew of 20 degrees). Then, they are classified in beams (which have headnotes), bar lines (longer than beams, without



(a) (b)

Fig. 2. (a) Original Image. (b) Graphical primitives detected.

headnotes and divide scores in bar units) and others (e.g. lines that are part of another kind of symbols).

Detection of filled headnotes is performed with a morphological opening (with a disk of *radius = w/3*, where w is the distance between staff lines) and using parameters of circularity, area and compactness.

Extraction of whole and half notes are more difficult because handwritten circles are often broken or incomplete, so morphological operations cause a lot of false positives and further work is required. After that, the remaining image is processed to obtain other graphic primitives.

Figure 2(a) shows the original and skewed image. Using Hough Transform the orientation of the image is detected and corrected. Thus, horizontal projections show every staff line as a maximum, and staff lines can be deleted. In Fig. 2(b) we can see the detection of graphical primitives: headnotes and vertical lines are in black color, and bar lines are shown as the thickest vertical lines. The remaining score is in grey color (staff lines are not actually present, but in this figure they are shown on purpose). As we can see, good results are achieved.

4 Old handwritten scores

A 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. Our work is 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 preservation and dissemination of artistic and cultural heritage). Working with old scores makes the task more difficult because of paper degradation (most scores are in poor condition) and the lack of a standard notation. In addition, there are distortions caused by staff lines (which are often handwritten), broken and touching symbols as well as high density of symbols. In order to cope with these problems, 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 the following sections, the method proposed to detect and extract staff lines and graphical primitives is exposed (see steps followed in Figure 3).

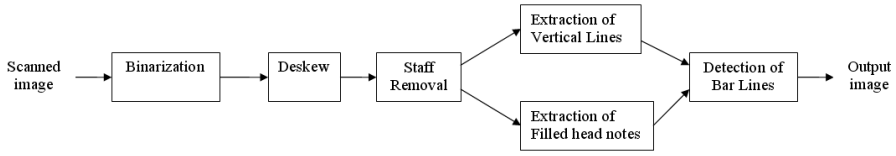


Fig. 3. Preprocessing Stages of the system.

4.1 Extraction of Staff lines

The preprocessing and segmentation phases must be adapted to this kind of scores: First of all, global binarization techniques do not work because of degradation of the scores; so adaptive binarization techniques are required (such as Niblack binarization [5]). Secondly, the detection of staff lines is more difficult due to distortions in staff (lines present often gaps in between), and contrary to modern scores, staff lines are rarely perfectly horizontal. This is caused by the degradation of old paper, the warping effect and the inherent distortion of handwritten strokes (staff lines are often drawn by hand). For those reasons, a more sophisticated process is followed (see Fig. 4).

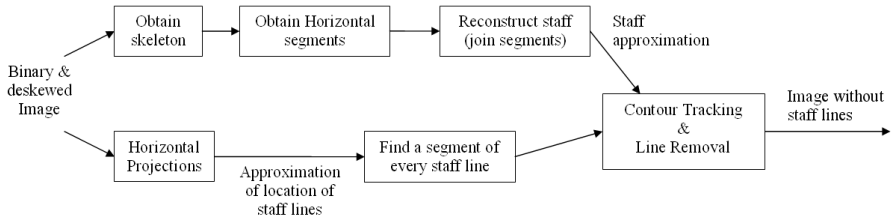
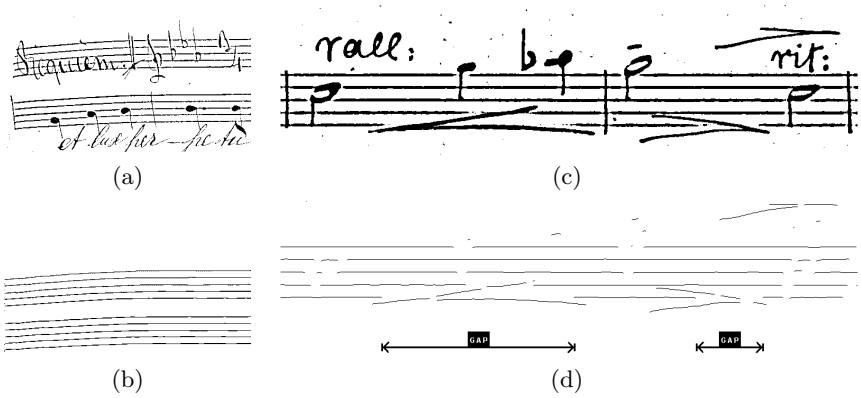


Fig. 4. Stages of the extraction of staff lines.

Here, Hough Transform is only used to deskew the input image, so horizontal projections can obtain a rough approximation of the location of staff lines. Then, horizontal runs are used as seeds to detect a segment of every staff line, and a contour tracking process is performed 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, a coarse staff approximation needs to be consulted.

The steps applied to obtain an image with horizontal segments (which will be candidates to form staff lines) are: First, the skeleton of the image is obtained and a median filter is applied with a horizontal mask. This process is repeated until stability (last two images are similar). In the output image, only staff lines and those horizontally-shaped symbols will remain. Notice that in a binary image, a median filter puts a pixel to 0 if most pixels of the neighborhood are 0, otherwise, its value will be 1. The size of this horizontal mask is constant (experimentally, dimensions are set to 1×9 pixels), because in the skeletonized image, each line is one pixel-width, so the width of lines in the original image is irrelevant.

In order to reconstruct the stave lines, each segment is discarded or joined with others according to its orientation, distance and area. Fig. 5(a) shows an original score suffering from a warping effect and its staff reconstruction (Fig. 5(b)). If there are big gaps in staff lines in presence of horizonal symbols this method could fail and follow a segment of this symbol instead of a segment of the staff line. Fig. 5(c) shows a big gap with a crescendo marking and Fig. 5(d) shows its reconstruction. An initial solution to it consists in increasing the size of the slice, but it could not work in scores with large deviations in staff lines.



(a) (c) (b) (d)

Fig. 5. (a) Original Image (b) Reconstruction of staff lines. (c) Original Image (d) Line segments of staff lines with gaps and horizontal symbols

Once we have obtained the reconstructed staff lines, the contour tracking process can be performed following the best fit path according to a given direction. If there is no presence of staff line (a gap), the 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 (depending on the width of staff lines, calculated using the statistical mode of line-segments) will be removed. Fig. 6 shows some examples of line removal: Fig. 6(b) is the original image, and in Fig. 6(a) we can

see how in presence of a gap, the process can detect next segment of staff line to continue; in Fig. 6(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. 6(d) and whose width is under this threshold, because context information is not still available.

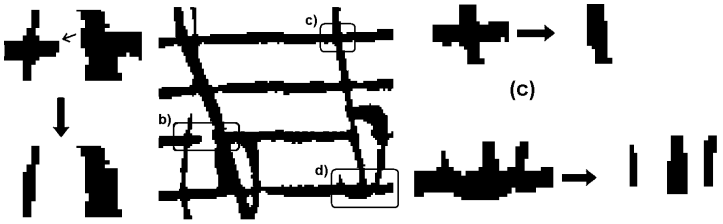


Fig. 6. Examples of Line Removal in Contour Tracking process. a) Gap in line, b) Original Image, c) Symbol crosses the staff line, d) Symbol is tangent to staff line: Symbol becomes broken

4.2 Recognition of vertical lines

After deleting staff and calculating the 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. 7(a)). 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 (as we do in modern scores), but results using median filters are better and the algorithm is faster.

4.3 Recognition of filled head notes

Working with printed scores makes this process easier, because all headnotes 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 degrees, discarding elements with large area. This approach gets all filled headnotes and false positives (Fig. 7(b)), but it is better to discard false positives in

next stages than forgetting real head notes. Because some modern rules of musical notation are not applied in old scores, we will classify notes (filled headnotes with beams) in higher-level stages, using grammar rules and the knowledge of time signature.

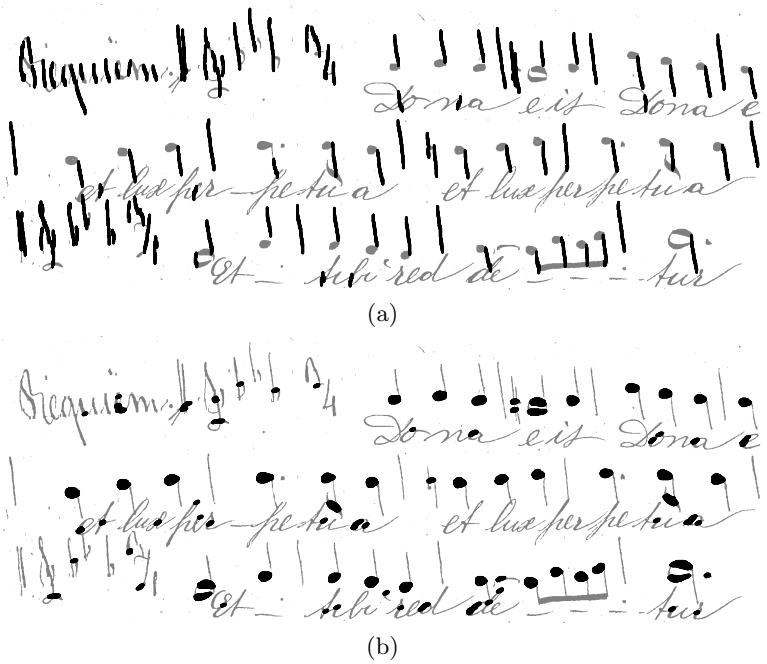


Fig. 7. A section of the Requiem Mass of the composer Aleix: (a) Vertical lines detected are in black color(b) Filled head notes detected in black color.

4.4 Recognition of bar lines

Once we have detected vertical lines and filled head notes, lines must be classified (see Fig. 8(a)) in beams (which have headnotes), bar lines (longer than beams, without headnotes) and others (e.g. lines that are part of another kind of symbols). 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.

A first approximation of bar lines is performed assuming that bar lines cover all staff and there are no headnotes in their extremes. So, if a vertical line is large enough and it is situated covering all five staff lines, then it will labelled as a bar line if there is no presence of filled headnotes in its extremes, see Fig. 8(b).

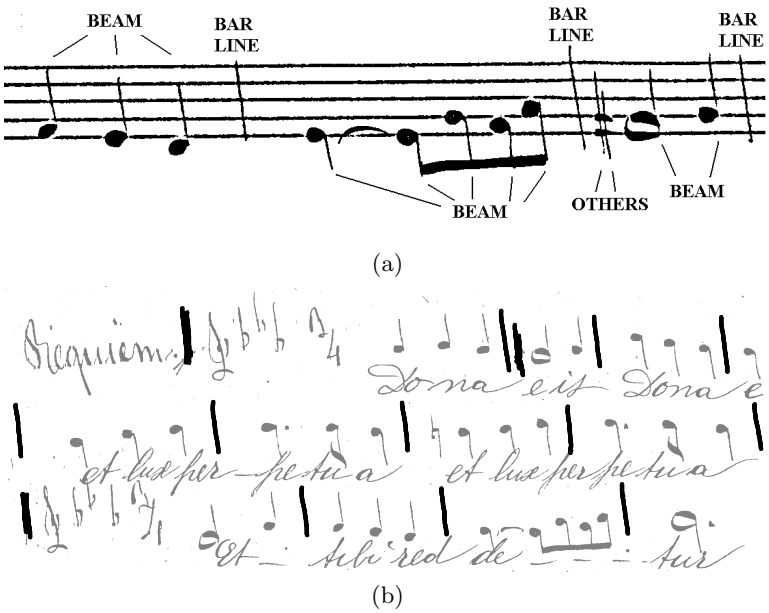


Fig. 8. (a) Verticals in scores(b) Bar lines in black color.

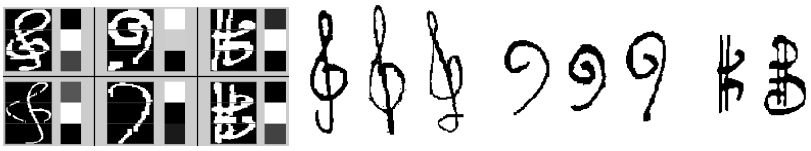
4.5 Classification of Clefs

Once every measure of the score is obtained, it is processed independently in order to recognize and classify all musical symbols. The heading of every score is formed of the clef, time signature and key signature. Because the clef determines the pitch of every note, it should be one of the first elements to recognize.

Due to the enormous variations in handwritten clefs, the classification of clefs must cope with deformations and variations in writing style. Thus, the method proposed uses Zernike moments (which maintain properties of the shape, being invariant in front of deformations) and Zoning, which codifies shapes based in statistical distribution of points in a compact and easy way. A full description of these techniques can be found in [7].

Zoning consists in computing the percentage of foreground pixels in each zone: an *mxn* grid is superimposed on the character image, and for each of the *nxm* zones, the average gray level is computed, giving a feature vector of length *nxm*. Thanks to the fact that in bass clefs the top of the clef has the bigger area, the Zoning algorithm can be used for a initial classification of bass clefs: The image is divided in 3 rows and 1 column, and the zoning vector (3x1) is filled with its normalized area. If the first row has the biggest area of the vector (see squares in white color in Fig. 9(a)), then the clef is a bass clef. Afterwards, clefs not classified with Zoning will be classified using Zernike Moments.

Zernike moments are defined over a set of complex polynomials which forms a complete orthogonal set over the unit disk.



(a) (b)

Fig. 9. (a) The application of Zoning technique to clefs using 3 files and 1 column to divide the images. (b) Clef Models for the classification using Zernike moments

Polynomials of Zernike are denoted by:

$$ZP = \{V_{nm}(x,y) | x^2 + y^2 \leq 1\} \tag{1}$$

The form of the Zernike polynomial basis of order n an repetition m ($n \in N^+$, $m \in N, (n - |m|)$ even, and $|m| \leq n$) and the radial polynomial are defined as:

$$V_{nm}(x,y) = R_{nm}(x,y) \exp(jm \arctan(y/x)); \tag{2}$$

$$R_{nm}(x,y) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s!(\frac{n+|m|}{2} - s)! \cdot (\frac{n-|m|}{2} - s)!} \cdot (x^2 + y^2)^{(n-2s)/2} \tag{3}$$

In our approach, 12 Zernike moments are used with 8 model classes for the three existing clefs (see Fig. 9(b)). The method normalizes the image of every model of the class and computes the Zernike moments and the feature vector. Afterwards, the Zernike moments and feature vector of the clef to be identified are computed. Then, the method will associate the new clef with the model class whose feature vector is closer to the feature vector of the clef to be classified.

5 Results

We have tested our method with a set of scores from the XIX century of several composers. These images of scores have been obtained through the archive of Seminar of Barcelona. Referring the staff removal stage, several pages of scores from different composers have been tested. In table 1 we can see that most staff lines are perfectly reconstructed, but sometimes (see section 4.1), a horizontal symbol is drawn over a staff line and causes the staff reconstruction to follow wrongly this symbol.

Concerning detection of graphical primitives, several staffs of scores from different composers have been tested (see an example in Figure 10). In table 2 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). More exhaustive results can be found in [8]. Performance in detection of filled head notes decreases when strokes are very thick, so in such cases, other objects could also be detected

Page	N. staffs	Perfectly Reconstructed / Total, (%)	Perfectly Removed / Total, (%)
1	10	49 / 50 , 98%	48 / 50 , 96%
2	10	50 / 50 , 100%	50 / 50 , 100%
3	10	45 / 50 , 90%	45 / 50 , 90%
4	10	49 / 50 , 98%	48 / 50 , 90%
5	12	54 / 60 , 90%	53 / 60 , 88%
6	14	70 / 70 , 100%	70 / 70 , 100%
7	14	69 / 70 , 98%	69 / 70 , 98%

Table 1. Staff removal results: When lines are not perfectly reconstructed, it is impossible to reach rates of 100% in staff removal

Page	N.Staffs	Verticals: C/D, (%FP)	Bar lines, (%FP)	Head notes, (%FP)
1	10	236 / 352 , 33%	71 / 80 , 11%	99 / 462 , 78%
2	10	177 / 237 , 25%	54 / 57 , 5%	96 / 465 , 79%
3	7	225 / 269 , 16%	40 / 43 , 7%	135 / 382 , 64%
4	7	218 / 284 , 23%	48 / 49 , 2%	128 / 365 , 65%
5	6	227 / 271 , 16%	38 / 41 , 7%	110 / 390 , 71%
6	6	180 / 254 , 29%	37 / 48 , 23%	122 /435 , 72%

Table 2. Results: 100% of Head notes, Vertical and Bar lines detected. Correct/Detected and FP= % of False Positives

as filled headnotes. Although there are many false positives, it is better to discard them in next stages than having false negatives (filled headnotes in thin strokes not detected). Finally, the classification of clefs reaches rates of 86% (44 clefs correctly described of 55 existing clefs).

6 Conclusions

In this work an approach to segment primitive elements in handwritten old music scores has been presented. Our strategy consisted of the following steps: First, score line detection and removal, using Hough Transform and a line tracking algorithm. Then, the detection of vertical lines and circular primitives is performed. Finally, the classification of vertical lines and clefs is described.

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 extracting lyrics from the scores, improving the reconstruction of staff lines, obtaining other graphic primitives and formalizing a grammar 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.

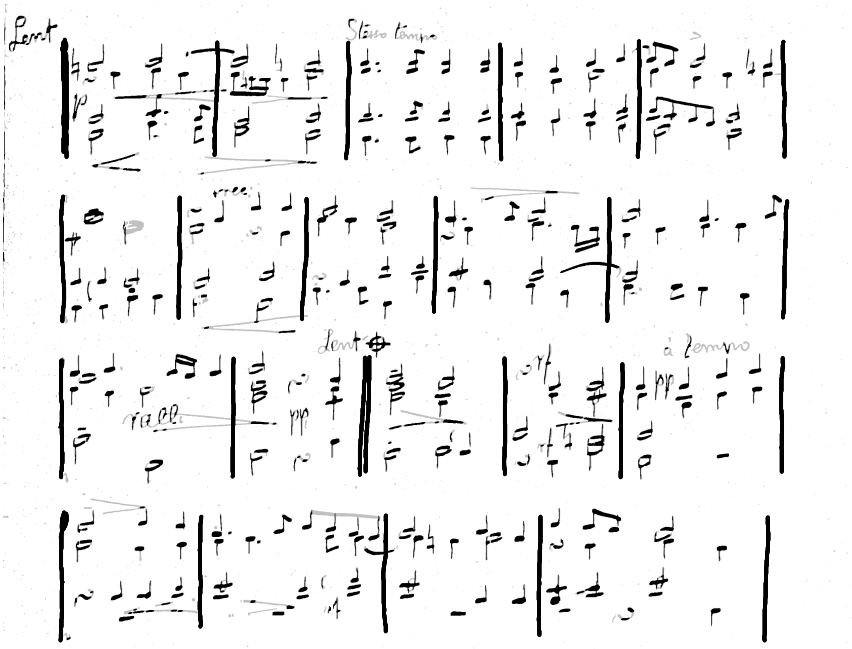


Fig. 10. Results from a section of "Salve Regina" of the composer Aichinger: Filled headnotes and beams in black color. Bar lines are the thickest lines

References

1. D. Blostein, H. Baird, "A Critical Survey of Music Image Analysis," *Structured Document Image Analysis*, Eds. H. Baird, H. Bunke, and K. Yamamoto, Springer Verlag (1992), 405–434.
2. K.C. Ng, "Music Manuscript Tracing", *Proceedings of the Fourth IAPR International Workshop on Graphics Recognition (GREC)*, Kingston, Ontario, Canada (2001), 470–481.
3. J.C. Pinto, P. Vieira, J.M. Sosa, "A New Graph-like Classification Method Applied to Ancient Handwritten Musical Symbols", *International Journal of Document Analysis and Recognition (IJDAR)*, Vol. 6, Issue 1 (2003), 10–22.
4. H. Kato and S. Inokuchi, "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 (1990).
5. W. Niblack, *An Introduction to Digital Image Processing*, Englewood Cliffs, Prentice Hall (1986), 115–116.
6. D. Bainbridge, N. Carter, "Automatic Reading of Music Notation", *Handbook of Character Recognition and Document Image Analysis*, eds. H.Bunke and P.S.P.Wang, World Scientific, Singapore (1997), 583–603.
7. Ø. D. Trier, "Goal-directed Evaluation of Binarization Methods", in Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(12), (1995).
8. A. Fornés, "Analysis of Old Handwritten Musical Scores", Master's Thesis, Universitat Autònoma de Barcelona, Spain (2005).