Optical Music Sheet Segmentation

P. Bellini, I. Bruno, P. Nesi
Dep. of Systems and Informatic University of Florence
Via S. Marta, 3 50139 Florence, Italy
+39 055 4796425, +39 055 4796365, +39 055 4796523
pbellini@dsi.unifi.it, ivanb@dsi.unifi.it, nesi@dsi.unifi.it,

Abstract

The optical music recognition problem has been addressed in several manners obtaining suitable results only when simple music constructs are processed. The most critical phase of the optical music recognition process is the first analysis of the image sheet. The first analysis consists of segmenting the acquired sheet in smaller parts. These may be processed to recognize the basic symbols. In this paper, the segmentation module of O^{3}MR system (Object Oriented Optical Music Recognition) system is presented. The proposed approach is based on the adoption of projections for the extraction of basic symbols that constitute graphic element of the music notation. A set of examples is also included in the paper.

Keywords: music segmentation, image processing, optical music recognition.

1 Introduction

Systems for music score recognition are traditionally called *OMR (Optical Music Recognition);* this term is tightly linked to *OCR (Optical Character Recognition).* Typically, OCR techniques can not be used in music score recognition since music notation presents a two dimensional structure: in a staff the horizontal position denotes different duration for notes and the vertical position defines the height of the note [1]. Several symbols are placed along these two directions.

The OMR task is quite complex since several composite symbols are arranged around the note heads. Despite to the availability of several commercial OMRs: PIANOSCAN, NOTESCAN in Nightingale, MIDISCAN in FINALE, PhotoScore in Sibelius, SmartScore, SharpEye, etc., none of these is satisfactory in terms of precision and reliability. They provide a real efficiency close far from the 100% only when quite regular music sheets are processed. This justifies the research works around the definition of reliable *OMR* algorithms.

In papers concerning *OMR* (a survey from 1960s to 1997 is [1], [3], [4], [6], [7], [8], [9], [10], [11], [12]). The OMRs can be mainly classified on the basis of the granulation chosen to recognize symbols of the music score. There are two main approaches to define basic symbols. The basic symbols can be considered: (i) the connected components remaining after staff lines removal (chords, beams with notes, etc.), or (ii) the elementary graphic symbols such as note heads, rests, hooks, dots, that can be composed to build music notation [2], [5]. With the first approach the symbols can be easily isolated from the music sheet (segmented), however, the number of different symbols is very high. The architecture of an OMR system and the definition of basic symbols to be recognized are tightly related to the methods considered for symbol extraction/ segmentation and recognition. Most of the OMR processes present a first phase of music segmentation. This phase is the most critical of the entire process. *The segmentation of basic symbols has to be independent on the music score style, size and on the music processed, from simple (sequence of single notes) to complex (chords in beams with grace notes and several alterations, markers, slurs, expression, ornaments, etc.).*

A problem addressed in music score segmentation is the management of staff lines that touch the elementary symbols. The removal of overlapping lines requires a complex process of reconstruction of involved symbols (e.g., [4]), with corresponding loss of information. As a consequence some authors preferred to recognize symbols without removing staff lines.

In this paper, a segmentation method for extracting basic notation symbols is described. It is based on an extensive use of projection profiles for detecting and locating basic symbols without removing staff lines. This segmentation method is used in O^{3}MR system (Object Oriented Optical Music Recognition) at the place of the first version [5]. The most relevant improvement are (i) an automatic and flexible segmentation process based on topological features of music sheets; (ii) the adoption of new methods for the identification of basic symbols, and the extraction of additional information that is used in the recognition and reconstruction phases.

2 The General architecture

The goals of the O^{3}MR are: to cope with music sheets

Proceedings of the First International Conference on WEB Delivering of Music (WEDELMUSIC'01)
0-7695-1284-4/01 $17.00 © 2001 IEEE

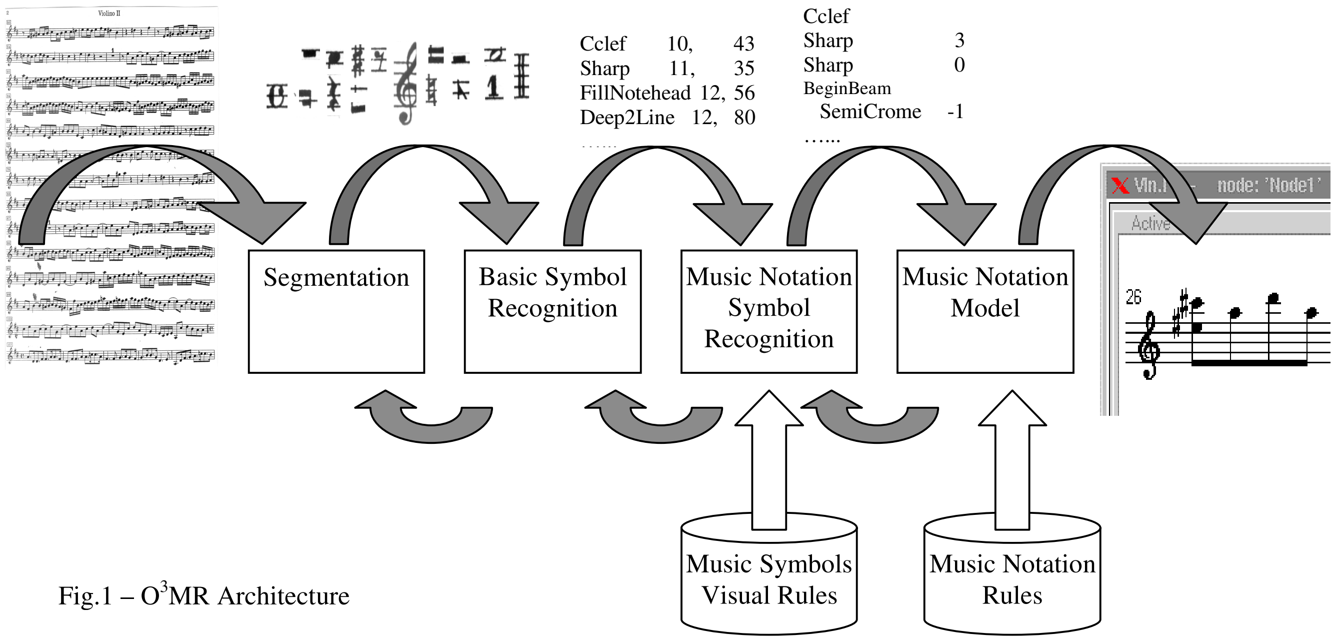

Fig.1 – O^{3}MR Architecture

written in the western music notation in which only monophonic music with five staff lines is present.

The general architecture of O^{3}MR is based on four main components (see Fig. 1):

Segmentation – the music sheet is processed with the aim of extracting the basic symbols (see Fig. 2) and their positions. From our point of view, the basic symbols are the elementary symbols that can be used for building the music notation symbols. For example: the filled note head; the deep lines representing beams, rests, sharps, flats; the empty note head; the thin lines used for drawing the staff, stem, slur, tie, wedges, etc. This means that each basic symbol can be used for building several music notation symbols. The exact identification is performed in the third block of O^{3}MR architecture. The definition of the set of basic symbols has been performed by considering the capability of (i) the segmentation algorithm in automatically extracting the symbols and (ii) the phase of recognition.

Basic Symbol Recognition module performs the recognition of basic symbols by using a neural network. It takes in input the normalized image segments of the basic symbols. On the basis of the set of basic symbols a feed–forward neural network has been set and trained to perform the recognition. The output of this module is mainly symbolic. For each recognized basic symbol, the image segment coordinates and the confidence value of recognition are produced. When barlines are recognized, the corresponding image segment is elaborated in order to estimate the position of staff lines. This information is extracted and communicated to the successive module to be considered as reference lines.

Music Notation Symbol Recognition. In this module, the recognized basic symbols are mapped into the elementary components of music notation symbols. For example, a deep line may be a part of a beam as well as of a rest, etc. The decision criteria are based on the recognition context: the position of the basic symbols with respect to the position of staff lines, the confidence level of the first phase of recognition, etc. In order to make simple the identification of elementary symbols, on the basis of their possible relationships, the **Visual Rules of the Music Symbols** have been formalized and used during the recognition. According to this process, for each basic symbol a set of probable elementary symbols are assigned. These elementary notation symbols estimate the probability to be basic symbols on the basis of the context. This module may request some additional evaluations when the decision cannot be taken with the current knowledge – for example when two Music Notation Symbols are similarly probable.

Music Notation Model Refinement – once the basic notation symbols are identified they are composed on the basis of a set of **Music Notation Rules**. The music model is reconstructed by refining the recognition performed in the previous phase. In this phase, the O^{3}MR is totally supported by MOODS music editor [5] in which the music rules are formalized and the correctness of each measure is easily checked.

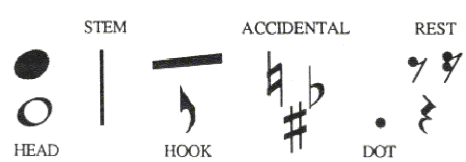

Fig.2 – Example of basic symbols (primitives)

In this paper, only the module addressing the

Proceedings of the First International Conference on WEB Delivering of Music (WEDELMUSIC'01)
0-7695-1284-4/01 $17.00 © 2001 IEEE

segmentation is discussed. To identify the most efficient solution, several experiments have been performed and a specific tool to perform the tests has been implemented. The same tool can be used for tuning and estimating the segmentation parameters as discussed in the following.

3 Segmentation

Music notation may present very complex constructs and several styles. Music notation symbols are various and can be combined in different manner to realize several and complex configurations [2]. This aspect impacts on the complexity of the segmentation problem.

An image of a music score page grabbed with a scanner is the starting point of the segmentation process. The segmentation method proposed is based on a hierarchical decomposition of the music image. The music sheet image is analyzed and recursively split into smaller blocks by defining a set of horizontal and vertical cut lines that allow isolating/extracting basic symbols (see Fig.3).

In more details, the procedure is based on the three elaboration levels depicted in Fig.4 and shortly commented as follows:

Level 0: the music sheet is segmented to extract sub images including the single music staff. In addition, a set of image score parameters are estimated for tuning the next processing phases.

Level 1: the image segment of each staff is processed to extract image subsegments that include music symbols and have a width close to that of note heads. This level is performed in three steps: (i) extraction of beams (e.g., group of beamed notes) and isolated symbols (e.g., clefs, rest, barline); (ii) detection and labeling of note heads; (iii) detection of other music symbols or parts of them.

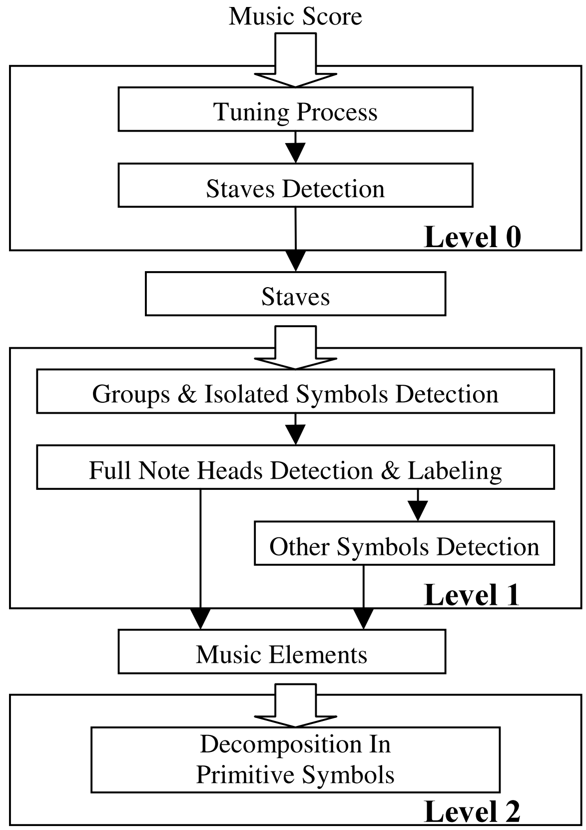

Fig.4 – Elaboration levels

Level 2: music symbols, detected at level 1, are decomposed in basic symbols. In this phase, two decomposition methods are used: for image segments containing note heads and for those in which they are missing. In this last case, the image segment may contain other symbols.

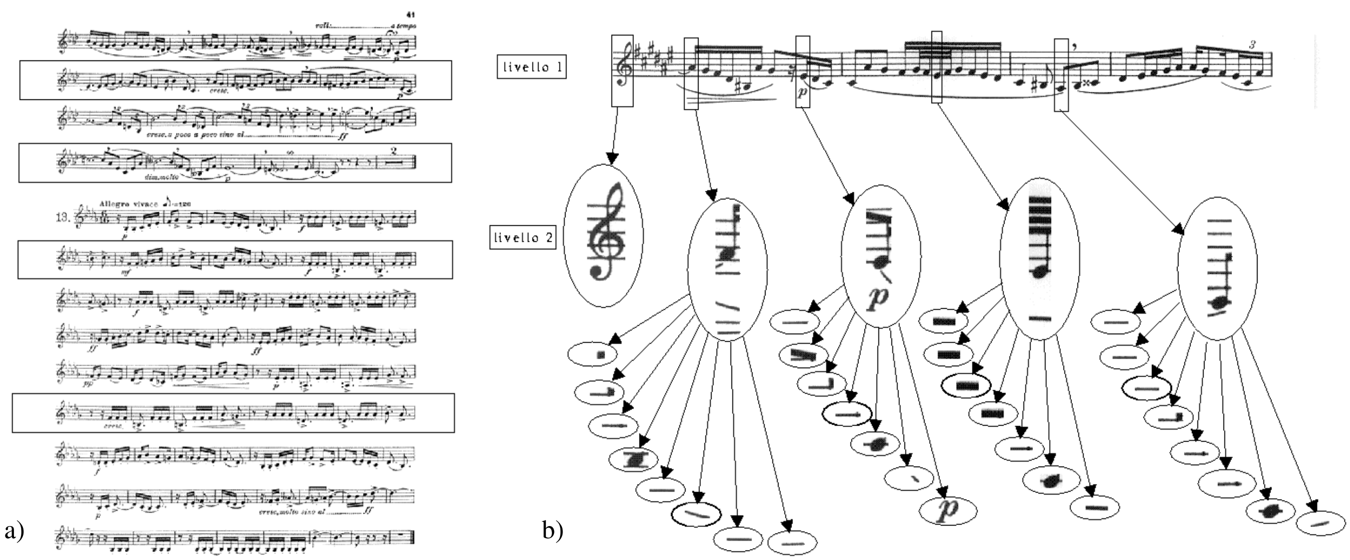

Fig.3 – a) Staff detection (Level 0); b) Decomposition in basic symbols (Levels 1 and 2).

Proceedings of the First International Conference on WEB Delivering of Music (WEDELMUSIC'01)
0-7695-1284-4/01 $17.00 © 2001 IEEE

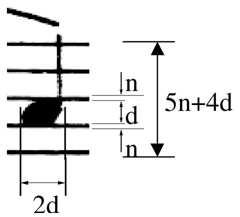

Fig.5 – Tuning

The segmentation procedure is based on the combined application of the X-Y projection profiles technique and image processing algorithms. This choice permitted to develop a set of algorithms which recursively operate on smaller image segments. This allows reducing the influence of the staff inclination or distortion.

Before to proceed in the detailed description of the above-mentioned levels, it should be noted that the images of the music sheets may have different magnitude. This implies that the dimension of music notation symbols and staves are unknown. This is a problem since the knowledge of the typical parameters allows better identifying the basic symbols. Typical formatting rules of music adopt well-defined ratio among the parameters depicted in Fig.5 and the size of all other music symbols [2].

Staff lines are graphic symbols that are independent on the music content. They give important information about music sheet features, since thickness of staff lines, n, and the distance between staff lines, d, are useful to tune the segmentation process, as shown in Fig.5.

For these reasons, differently from [1], [4], [7], [9], [10], [11] in our approach, staff lines are not removed since they are used in the segmentation process. This avoided the introduction of elaboration phases to fix symbols partially cancelled by the staff lines removal. The knowledge of staff lines position allows detecting the right pitch of notes in the reconstruction phase.

3.1 Level 0

The main goals of Level 0 are the (i) tuning of the segmentation process by the identification of a set of graphic parameters, (ii) detection of image segments in which staffs are included. According to the above-presented hierarchical structure, the image is decomposed in a set of sub images, each of which includes a staff.

Tuning Process -- The tuning process is performed to estimate the music sheet parameters from the scanned image. These parameters are (i) the thickness, *n,* of staff lines, and (ii) the distance *d* between staff lines (see Fig.5). To estimate these values, the score image is processed column by column to generate two histograms in which the size of sequences of black and white pixels are counted as a function of their dimension. In the histograms, the occurrence value and the number of pixels are considered; the number of pixels is positioned on X-axis while the number of occurrence value on Y-axis. These histograms represent the profile of runs for black and white pixels, respectively. As shown in Figs 6a and 6b, the most frequent values for *d* and *n* are the absolute maximum values of those curves, respectively. In order to manage variability and noise on the values of these parameters, two intervals have been adopted as follows:

- n_1 and n_2: the minimum and the maximum values for the staff line thickness, respectively
- d_1 and d_2: the minimum and the maximum values for the distance between two staff lines, respectively

These values are related to the maximum peak of each histogram. The histograms present a peak with a wideness that depends on the image acquisition resolution and on the original staff line thickness. Cutting the peak at 1/3 of its value leads to identify the measure of its wideness. In this view, the thickness of line is assumed to be [n_1, n_2], by taking the approximation for excess of the measure performed. In the same manner, it is obtained [d_1, d_2]. The above parameters are used in the next steps of the segmentation process.

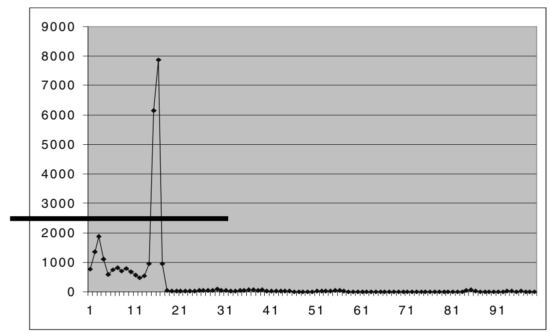

Fig. 6a – Distance between two staff lines, *d*

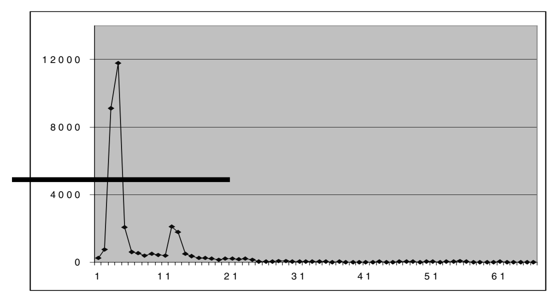

Fig. 6b – Thickness of staff line, *n*

Staves Detection and Isolation -- According to the hierarchical decomposition, the staff detection is the first elaboration phase of the O^{3}MR system. The goal is to identify a rectangular area in which the staff is located in order to process that image segment to extract the contained basic music symbols. The algorithm for detecting the staffs is based on the recognition of the staff line profile. The profile, obtained by applying the Y-projection to a portion of staff lines image, presents a regular pattern in terms of structure whereas other projections have a variable pattern. Please see Fig.7, in which the staff lines are zoomed and reported in black on the left. In order to distinguish the projection of lines from the other graphic elements, a transformation of profiles has been introduced.

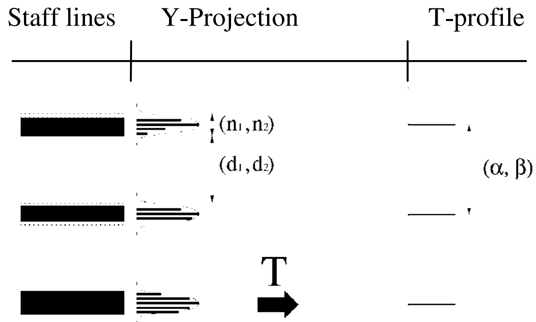

Fig.7 – T-transformation of staff lines profile

In details, the transformation, T, works on the Y-projection of a vertical image segment. The Y-projection is constituted by a set of groups/peaks, each of which is associated with a staff line. For each of them, their width is evaluated. When the width is comparable with values defined by [n_1, n_2], then the position of the mean position for the peak is estimated, otherwise it is not considered. The position of the mean values defines the lines in the T-Profile of Fig.7, and allows characterizing the staff in the identification phase. The analysis of the distance between lines in T-profile is used to understand if the profile is due to the presence of a staff.

The distance between lines of the T-domain is strictly related to the values of the above mentioned parameters. In fact, given the [n_1, n_2] range for the thickness of staff lines and the [d_1, d_2] range for the distance between two lines, the distance between mean values expressed in term of interval [α, β] is defined as: $\alpha = d_1$ and $\beta = d_2 + 2(n_2-(n_2-n_1)/2)$.

The staff detection algorithm looks for the "five equidistant" lines structure. This is performed by analyzing thin slices of the image sheet. Each slice has a vertical size equal to the whole image score size and width equal to a few pixels (dx). The slices processed performing T-transformation. On each slice (see Fig. 8), a probe to look for the five lines pattern is used. The probe is applied on a sub-section of the slice and analyzes it from top to bottom. The probe looks for the starting coordinate of the 5 lines staff. For this reason, its height, I, has to be comparable with staff height. The I value is defined as a function of the thickness parameters considering the maximum values of n and d ranges: $I = 5n_2+4d_2$

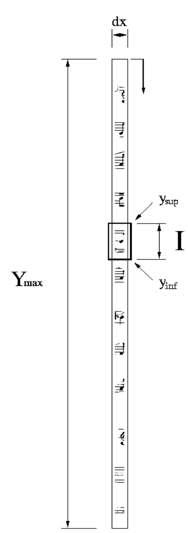

Fig.8 – Probe window

In order to cope with eventual staff deformations, the above value of I has been increased of 20%. The staff detection by means of the probe is performed in two phases: (i) discovering and (ii) centering the staff. In the discovering phase, the probe window is used to detect the staff lines and store the starting coordinates of segment in which the staffs are present. In the centering phase, a couple of staff coordinates (y_{sup}, y_{inf}) are obtained. These are defined in order to fix the cut lines for extracting the staffs contained in the score image. If n is the number of staffs, then each new couple of coordinates has been evaluated as:

$$\begin{cases} \hat{y}_{\text{sup}}^{(1)} = 0 \\ \hat{y}_{\text{inf}}^{(1)} = \dfrac{y_{\text{sup}}^{(2)} + y_{\text{inf}}^{(1)}}{2} + \varepsilon \end{cases} \qquad \begin{cases} \hat{y}_{\text{sup}}^{(n)} = \dfrac{y_{\text{sup}}^{(n)} + y_{\text{inf}}^{(n-1)}}{2} - \varepsilon \\ \hat{y}_{\text{inf}}^{(n)} = Y\max \end{cases}$$

Where: $i = 2, \ldots, n-1$, $\varepsilon \geq 0$ (a tolerance value to increase robustness). The introduction of ε tolerance allows getting adjacent or partially overlapped image segments containing the staff. In the above definition, the coordinates of the first and last staff are excluded.

3.2 Level 1

Level 1 works on the image segments produced from Level 0, containing one staff. The aim of this level is to extract the vertical image segments containing music symbols. According to Fig.5, Level 1 produces the lower level sub-image segments in three phases for detection: (i) of groups and isolated symbols; (ii) and labeling of note heads; (iii) of other music symbols or parts of them. In certain cases, the third phase is not needed.

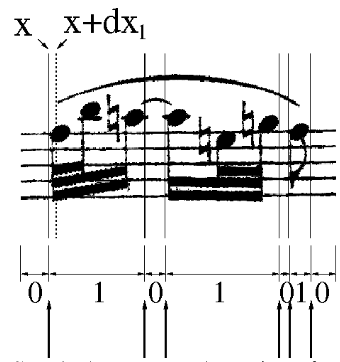

Fig.9 – Symbol segment detection, function F

Groups & isolated symbols detection -- In this phase, groups of figures and single symbols (e.g., clefs, rest, barline) are detected. Image segments in which the staff does not present symbols separate them. To this end, the staff detection algorithm is applied to produce the value of the binary function, F. The detection process has been realized by considering a running image window. This

Proceedings of the First International Conference on WEB Delivering of Music (WEDELMUSIC'01)
0-7695-1284-4/01 $17.00 © 2001 IEEE

has the height of the image segment coming from level 0, and width, dx_1, equal to 3-4 pixels. The analysis of the image segment is performed by moving the running window from left to right of one pixel a time. The result of staff detection sets the values of binary function, F (see Fig. 9). The 0 value is associated with the presence of an empty staff and 1 otherwise. In this phase, the stand-alone music symbols (clef, barline, time signature, whole notes, etc.) are detected. Whereas, non-stand alone music symbols and all groups of figures have to be processed in the next phase in order to proceed at their decomposition in smaller image segments.

The identification of empty staff allows processing of the corresponding image segment in order to estimate the coordinates of staff lines. This information is used in Level 2 as described in the following.

Note head detection & labeling -- The goal of this phase is to slice the complex image segments produced by the previous phase and marked as F=1. In the slices produced by this phase one or mode single note heads have to be present along the y-axes. The proposed approach is based on searching the presence of single note head. In western notation, note heads may present at least a diameter equal to the distance between two staff lines. To consider the note head width equal to $2d_1$ is a reasonable approximation. For these reasons, image segments coming from the previous phase are processed on the basis of their width. In this case, only image segments larger than $2d_1$ are considered. In the X-projection, we have: (i) spikes due to note stems and vertical symbols; (ii) offsets due to horizontal symbols like staff lines, beams, slurs, crescendo, etc. (iii) smoothed dense profile due to note head.

In order to extract the note heads the dense profile contribution has to be extracted from the X-projection. This means to eliminate the other two contributions.

To this end, a thin running window is considered on the image segment containing the staff with symbols (see Fig.10). The running window scans the image with a step equal to 1 pixel. For each step/pixel the Y-projection is calculated. In the projection, the presence of a note head can be detected on the basis of its width, H, which is in the range $[2n_2, 2n_2+d_1]$. Since the objective of the process is to detect the location of note heads, only the maximum value of H along the projection of the running window is considered (Hy6). This value is reported in the final X-projection shown in Figs.10 and 11a.

The note heads produce higher peaks since they are deeper than beam lines. On the other hand, when note heads are missing the maximum value produced from the beam is reported in the X-projection of max Y-projection. The evident difference in the value of the two cases is amplified by the adoption of a running window that brings to consider several overlapped thin windows of the same note head.

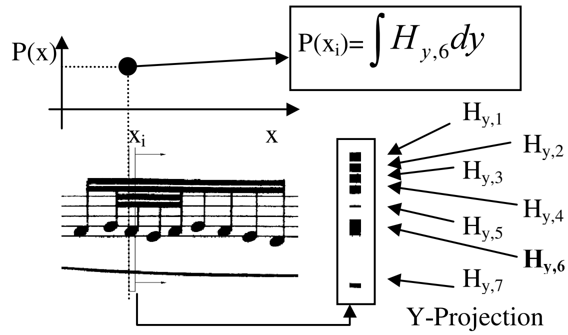

Fig.10 – X-Projection building

The final step consists of isolating the note heads. This is performed by using a double threshold mechanism to obtain the results shown in Fig.11b. The first threshold is defined as the compromise between the dimension of the note head and that of the beams. The second filtering is performed on the width of the remaining peaks that are considered as due to the presence of note only if their width is larger than $d_1/2$. For each peak of the X-projection the mean position of the peak along the X-projection is estimated. If C is the x coordinate of the mean value, the couple of points is defined as following: $(C-d_1, C+d_1)$.

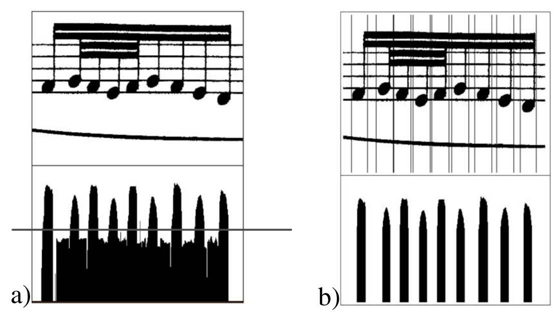

Fig. 11 – a) X-projection before thresholds application; b) X-projection after thresholds application and ready for extracting image segments with note heads.

Each couple of points defines the width of the image segment that includes the note. In particular, the width is equivalent to that of the note head (see Fig.11a).

In this process, each image segment containing a note head is labeled. This information is used by Level 2. The above process for labeling segments that contains note heads is also performed on the image segments marked with F=1.

Please note that with the presented approach also chords having notes on the same side are managed (see next section). Image segments that do not contain note heads are non-sliced by the described process.

Other Symbols Detection -- The objective of this step is to analyze the image segments of the staff that have been excluded in the previous steps. Image segments that have to be processed typically include groups of symbols very close to each other. In most cases, these symbols are

Proceedings of the First International Conference on WEB Delivering of Music (WEDELMUSIC'01)
0-7695-1284-4/01 $17.00 © 2001 IEEE

separated by small spaces generating local minima in the X-projection profile, such as in Fig.12. Thus, detecting these points means to allow slicing the image segment and thus to extract the basic symbols. To this end, an iterative method has been developed. As a first step, a low pass filter is applied to smooth the profile. The smoothed profile is analyzed in order to find the position of the minima. Its position is used to divide the image segment in two new sub-images. The same procedure is applied at the two sub-images when their width is greater or equal than $d_1/2$. The process stops when the maximum width of the processed image segments is lower than $(5/2)\ d_1$. The last operation consists in sorting the points for vertical image segment in order to define image segments.

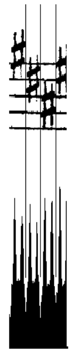

Fig.12 – Isolation points and minimums of X-projection

In presence of a "constant" profile the segmentation process produces image segments with a width comparable to that of note heads, $d_1/2$. Image segments having a width lower than the staff line thickness are not passed to the next segmentation level. This process is capable to cope with key signatures, and accidentals, since it allows decomposing the image segments non-decomposed in the previous step.

3.3 Level 2

Level 2 is the last phase of the segmentation process. In this phase the images coming from the previous Level 1 are decomposed into a set of basic symbols. This phase covers an important role since the recognition and reconstruction are strictly connected to it. The produced image segments must include graphic details: (i) belonging to the set of defined basic symbols in repeatable manner, (ii) additional information needed for their recognition. The first aspect impacts on the reliability of the recognition phase, while the second on the application of the rules of the reconstruction phase.

The previous segmentation phase produced different results depending on the presence or not of note head. In this process, two different segmentation methods are applied to the received image segments: (i) including, and (ii) non-including note heads. This distinction is possible by using the information given by labels defined in the note head detection phase. Both segmentation methods are based on the Y/X-projection and produce couples of vertical coordinates to extract basic symbols.

Images with note heads -- In image segments containing a note head, this can be connected to other basic symbols. This implies that a specific process for their division is needed. Ornaments (turn, mordent, trill, etc.) and horizontal (slurs, crescendo, etc.) symbols can be more easy identified since they are not connected or strongly close to the note head. The graphic components of the note are mainly: (i) note head, and (ii) beams or hooks. Observing the Y-projection of notes (see Fig.12a), it is possible to identify the contribution of the stem. It adds an offset to the profile of the projection linking the profile of the note head with beams and hooks. In the case of a quarter and whole note, it is only an offset. In both cases, the contribution of stem is used as the lower value for the threshold that is applied for the extraction of basic symbols. The proposed approach is based on the result obtained in [6] and it is comprised of the following steps as depicted in Fig.13:

1. **Staff lines removal from Y-projection.** According to the processing of the last segment containing an empty staff the position of the staff lines is known (see Level 1). The contribution of the staff lines is removed by masking their contribution lines having a width of n_2. Fig.13c shows the obtained Y-projection.
2. **High-pass filtering.** This filtering phase allows eliminating the low frequency components. Fig.13d shows the filtered Y-projection: the offset due to the stem has been removed.
3. **Extraction points computation**: it is based on the computation of extraction points by means of a threshold mechanism [7]. When two successive segments are closer than n_2 they are fused in unique segment (see Fig.13e).

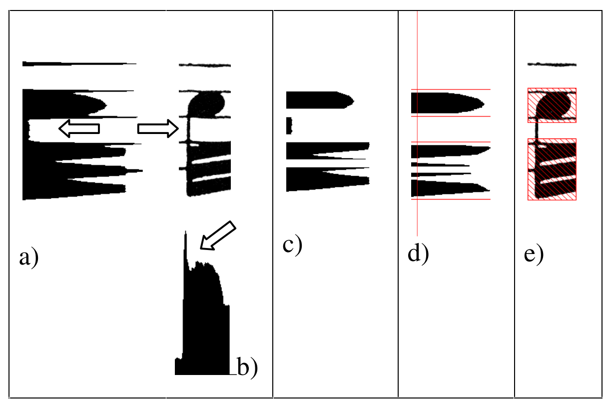

Fig. 13 – a) Y-projection of 32-th with beams. Arrows indicate the offset due to the stem; b) X-projection: the arrow indicates the spike due to the stem; c) Y-projection of the image without staff lines; d) Y-projection after staff lines removal, filtering and extraction points computation; e) identification of basic symbol after extraction point computation

4 More Complete Examples

In this section, some complete examples are presented. The image used represents a monophonic music sheet acquired at 300 dpi. The tuning phase

Proceedings of the First International Conference on WEB Delivering of Music (WEDELMUSIC'01)
0-7695-1284-4/01 $17.00 © 2001 IEEE

detected the following parameters: $[n_1, n_2] = [2, 4]$, $[d_1, d_2] = [17,18]$ pixels.

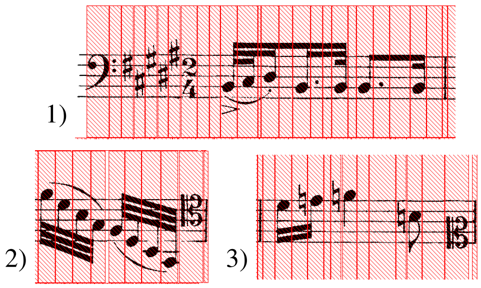

Fig. 14 – Music score, Image segments slicing

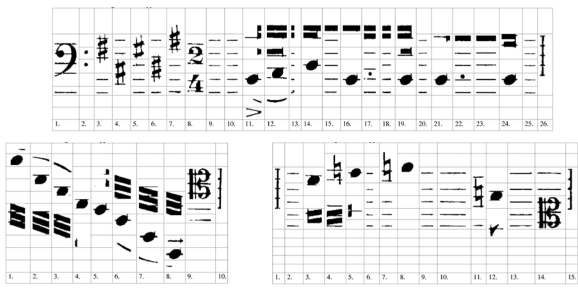

Fig. 15 – Decomposition in basic symbols

In Fig.14, the results of Level 1 for three measures are presented. Measures have been decomposed in vertical image segments and the clef has been completely included in an image as well as sharps. In the segmentation, each symbol has been correctly isolated. Fig.15 shows the last decomposition produced by Level 2. Basic symbols have been distributed per column to depict the decomposition of each vertical image segment coming from Level 1. According to Level 2 algorithms, image segments that include the note head and other symbols has been decomposed in different manner.

Finally, the segmentation method extracts beaming both in horizontal and in slanting direction. Barlines are completely extracted. In Fig.15 (third table), the eighth note has been decomposed in two main basic symbols: head and hook.

5 Conclusions

The optical music recognition is a key problem for coding music sheets. The most critical phase of the optical music recognition process is the first analysis of the image sheet. In optical processing of music or documents, the first analysis consists of segmenting the acquired page in smaller parts on which the recognition may be applied. The proposed segmentation method has produced interesting results considering both monophonic music score and single voice, and music for piano solo. It is currently the first module in the O^3MR architecture described in this paper and has strongly improved the general OMR process with respect to the previous version [6] (based on filtering). The tuning phase allows considering images coming from different publishers and acquired by the scanner at different resolutions. On the basis of the parameters fixed by the tuning phase, the staff detection algorithm is capable to cope with problems of the staff inclination, deformation and indentation. The note head identification and the other symbols detection algorithms are able to work with image presenting a high density of note per measure. The last segmentation level extracts basic symbols. The experimental results have shown a high efficiency in the correct location of basic symbols. Finally, this segmentation method produces basic symbols that can be collected in 54 categories. These are categories used by the neural network in the recognition process.

REFERENCES

[1] M. Roth, *An Approach to Recognition of Printed Music*, tech. rep., Swiss Federal Institute of Technology, ETH Zurich, Switzerland, 1994.

[2] T. Ross, *The Art of Music Engraving and Processing*, Hansen Books, Miami, 1970.

[3] D. Blostein and H. S. Baird, *A Critical Survey of Music Image Analysis*, in H. S. Baird, H. Bunke, and K. Yamamoto, editors, Structured Document Image Analysis, pp. 405-434, Springer, 1992.

[4] H. Kato and S. Inokuchi, *The Recognition System for Printed Piano Music Using Knowledge and Constraints*, in Proc. Workshop SSPR, pp.231-248, 1990.

[5] P. Bellini, F. Fioravanti and P. Nesi, *Managing Music in Orchestras*, IEEE Computer, Sept., 1999.

[6] S. Marinai, P. Nesi, *Projection Based Segmentation of Musical Sheet* Proc. of the 5th Intern. Conference on Document Analysis and Recognition, ICDAR'99, IEEE press, IAPR (International Association on Pattern Recognition), Bangalore, India, pp.515-518, 20-22 Settembre 1999.

[7] D. Bainbridge, *An Extensible Optical Music Recognition System*, in the Australasian Computer Science Conference (Melbourne 1996), pp. 308-317.

[8] N. P. Carter, *Automatic Recognition of Printed Music in Context Electronic Publishing*, Ph.D thesis, University of Surrey, February 1989.

[9] T. Kobayakawa, *Auto Music Score Recognizing System*, in Donald P. D'Amato editor, Proceedings SPIE, volume 1906, May 1993.

[10] A. Tojo, H. Aoyama, *Automatic Recognition of Music Score*, Proceedings of 6th International Conference on Pattern Recognition, DE (1982)

[11] I. Fujinaga, *Adaptive Optical Music Recognition*, Ph.D Dissertion. McGill University, Montreal, CA, 1997.

[12] E. Selfridge-Field, "Optical Recognition of Musical Notation: A Survey of Current Work," Computing in Musicology, Vol. 9, 1993-4, pp. 109-145.

Proceedings of the First International Conference on WEB Delivering of Music (WEDELMUSIC'01)
0-7695-1284-4/01 $17.00 © 2001 IEEE