Recognition of handwritten music scores

Arnau Baró-Mas

Resum– El reconeixement de partitures musicals manuscrites segueix sent un problema obert. Els enfocaments existents només poden reconéixer partitures manuscrites molt simples, principalment a causa de la variabilitat en l'estil d'escriptura i la variabilitat en la composició dels grups de notes musicals (p.e. els símbols musicals compostos). En aquest treball, per començar, se separaran els símbols simples (p.e blanques, negres, claus, sostinguts) dels compostos i els estudiaré per separat. Els símbols simples mitjançant mètodes de reconeixement de símbols i els compostos a partir d'una jerarquia de primitives i regles sintàctiques. El meu mètode ha estat provat utilitzant diferents partitures de música escrita a mà de la base de dades CVC-MUSCIMA i comparat amb un programari de reconeixement òptic musical comercial. Tenint en compte que el meu mètode és d'aprenentatge lliure, els resultats obtinguts són prometedors.

Paraules clau– Reconeixedor musical òptic; Partitures musicals manuscrites; Reconeixedor de símbols manuscrits; Agrupació Perceptual

Abstract– The recognition of handwritten music scores still remains an open problem. The existing approaches can only deal with very simple handwritten scores mainly because of the variability in the handwriting style and the variability in the composition of groups of music notes (i.e. compound music notes). In this work on the one hand I study the isolated symbols (i.e half-note, quarter-note, clefs, sharps) and on the other hand the compound music notes. Firstly, I will separate the isolated symbols (i.e half-notes, quarter-notes, clefs, sharps) to the compounds and I will study each one separately. The isolated symbols will be recognized with symbol recognition methods and compounds with a primitive hierarchy and syntactic rules. The method has been tested using several handwritten music scores of the CVC-MUSCIMA database and compared with a commercial Optical Music Recognition software. Given that my method is learning-free, the obtained results are promising.

Keywords– Optical Music Recognition; Handwritten Music Scores; Hand-drawn Symbol Recognition; Perceptual Grouping

1 INTRODUCTION

The recognition of music scores [1, 2, 3] has attracted the interest of the research community for decades. Since the first works in the 60s [4] and 70s [5], the recognition or music scores has significatively improved. In the case of printed music scores, one could say that the state of the art has reached a quite mature state. As a matter of fact, many commercial OMR systems show very good performance, such as PhotoScore [6] or SharpEye [7].

Nowadays, a lot of handwritten music scores exist only in the form of original or photocopied manuscripts, without being published in a digital format. Digitization is an essential tool to preserve this cultural heritage. For this, Optical Music Recognition (OMR) algorithms are required.

An OMR system has a previous image preprocessing stage (application of techniques such as enhancement, binarization, noise removal, among others). A typical OMR framework has three principal modules [8]: (1) recognition of musical symbols. The staff lines have to be removed and the primitives detected. (2) Reconstruction of the musical information. Here its common use graphical and syntactic rules to overcome possible classification errors and (3) construction of a musical notation model for its representation which can be a MIDI, digital music score, etc.

Concerning handwritten scores, although it is remarkable the work in Early musical notation [9, 10], the recognition of handwritten Western Musical Notation still remains a challenge. The main two reasons are the following. First, the high variability in the handwriting style increases the difficulties in the recognition of music symbols. Second,

• E-mail de contacte:arnau.baro@e-campus.uab.cat
• Menció realitzada: Computació
• Treball tutoritzat per: Alicia Fornés-Bisquerra (CC. de la Computació)
• Curs 2015/16

the music notation rules for creating compound music notes (groups of music notes) allow a high variability in appearance that require special attention.

In order to cope with the handwriting style variability when recognizing individual music symbols (e.g. clefs, accidentals, isolated notes), the community has used specific symbol recognition methods [11, 12] and learning-based techniques such as SVMs, HMMs or ANNs [13]. As stated in [14], in the case of the recognition of compound music notes, one must deal not only with the compositional music rules, but also with the ambiguities in the detection and classification of graphical primitives (e.g. headnotes, beams, stems, flags, etc.). It is true that temporal information is undoubtedly helpful in on-line music recognition, as it has been shown in [15, 16]. Nowadays, a musician can find several applications for mobile devices, such as StaffPad [17], MyScript Music [18] or NotateMe [19].

Concerning the off-line recognition of handwritten groups of music scores, much more research is still needed. PhotoScore seems to be the only software able to recognize off-line handwritten music scores, and its performance when recognizing groups of notes is still far from satisfactory. One of the main problems is probably the lack of sufficient training data for learning the high variability in the creation of groups of notes.

For these reasons, this work is focused on the off-line recognition of handwritten music scores, including both isolated symbols and compound music notes. For this purpose the music symbols will be first classified into isolated or compound. The isolated music symbols will be classified using a symbol recognition method, such as the Blurred Shape Model. For the recognition of compound music notes, the aim is to avoid the need of training data. Thus, I propose a learning-free hierarchical method inspired in perceptual grouping techniques that have been applied to text detection [20] and object recognition [21]. Basically, the idea is to hierarchically represent the graphical primitives according to perceptual grouping rules, and then, validate the groupings using music rules.

The rest of the paper is organized as follows. Firstly, the challenges in the recognition of music scores is described in Section 2. Motivation, objectives, methodology and temporal planning are described in Sections 3, 4, 5 and 6. Afterwards, Section 7 describes the preprocessing and the clasification of symbols into isolated/compound is described in Section 8. Next, Section 9 defines the method to the Isolated music notes. Section 10 explains the hierarchical representation to combine the graphics primitives into more complex elements, and the validation of each group hypothesis. Section 11 discusses the experimental results. Finally, conclusions and future work are drawn in Section 12.

2 Challenges in the recognition of music scores

Music scores are a particular kind of graphical document that include text and graphics. The graphical information corresponds to staffs, notes, rests, clefs, accidentals, etc., whereas textual information corresponds to dynamics, tempo markings, lyrics, etc. For this reason, layout analysis and text/graphics separation methods are also necessary.

Concerning the recognition of graphical information, Optical Music Recognition (OMR) has many similarities with Optical Character Recognition (OCR). In the case of the recognition of isolated music symbols (e.g. clefs, accidentals, rests, isolated music notes), the task is similar to the recognition of handwritten characters, digits or symbols. In this sense, the recognizer must deal with the variability in shape, size and visual appearance. Figure 1 shows several examples of diferents clefs.

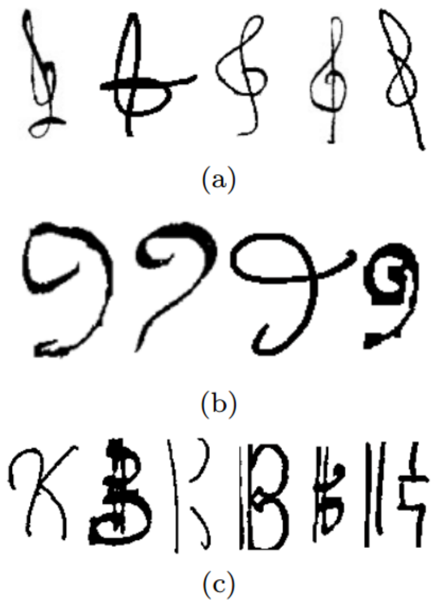


Fig. 1: High variability of hand drawn musical clefs: (a) Treble. (b) Bass. (c) Alto. Image extracted from [22]

Similarly, the recognition of compound music notes (i.e. notes joined using beams) could be seen as the task of recognizing handwritten words. It is nevertheless true that the difficulties in OMR are higher than in OCR because OMR requires the understanding of two-dimensional relationships, given that music elements are two-dimensional shapes. Indeed, music scores use a particular diagrammatic notation that follow the 2D structural rules defined by music theory. Music notation allows a huge freedom when connecting music notes, which increases the difficulties in the recognition and interpretation of compound notes. For example, music notes can connect horizontally (with beams), and vertically (chords), and the position and appearance highly depends on the pitch (melody), rhythm and the musical effects that the composer has in mind. Figure 2 shows several examples of compound music groups that are equivalent.

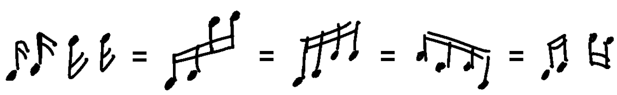


Fig. 2: Examples of equivalent Sixteenth notes.

Following the comparison with handwritten text recognition, it is true that language models can also be defined to improve the OMR results, just like language models and dictionaries help in handwriting recognition. For example, syntactical rules and grammars could be easily defined to cope with the ambiguities in the rhythm. In music theory, the time signature defines the amount of beats per bar unit. Therefore, all the music notes inside a bar unit must sum up to the defined amount of beats. Although grammars and rules [23, 24] have shown to be very useful to solve ambiguities, it is extremely difficult to use them when there are several melodic voices and chords, such as in polyphonic music. Moreover, it must be said that music *tuplets* (defined as irrational rhythms or extra-metric groupings) and ornament notes (e.g. *Appoggiatura*) scape from the beats restriction.

Finally, music semantics could also be defined using knowledge modeling techniques (e.g. ontologies), document understanding. Indeed, a musicologist could define the harmonic rules that should be applied for dealing with melodic ambiguities in polyphonic scores. However, these harmonic rules highly depend on the composer and the time period (e.g. some kinds of dissonant chords or intervals are only common in modern ages). Therefore, the incorporation of this knowledge seems unfeasible in this OMR stage.

3 Motivation

Firstly, when I started reading the different papers of recognition on the handwritten music scores I realized that there are few optical music recognition software and I was encouraged to make my own.

Additionally, I was interested in carrying out a final project on computer vision because it is a very fascinating topic. I was very enthusiastic to realize a final degree project mixing the computer vision with music. Finally, choosing this project I had the opportunity to write a paper for a conference. This point seemed me a great idea to finish off the project. On May 2016, I was able to send a paper for the International Conference on Frontiers in Handwriting Recognition (ICFHR-2016). The paper [25] is on the recognition of handwritten compound musical symbols groups of (eighth note, sixteenth note, etc).

4 Objectives

The aim of this project is a system to recognize handwritten music scores. I will study several techniques for image processing and classification. To achieve my goal, these are the subgoals:

- Study the current state of the art on handwritten music scores.
- Design a methodology for recognizing the handwritten music scores.
- Develop a classifier of Isolated and Compound symbols.
- Develop a symbol recognition method for isolated music notes and a perceptual grouping for compound music notes.
- Creating a ground-truth and validating the methodology.

5 Methodology

5.1 Software Development Framework

To develop this project I have chosen the Feature Driven Development (Based development features) as a methodology. With this agile methodology I am able to develop quality software with constant monitoring. FDD focuses on short iterations, allowing the delivery of tangible product in a short period of time, at the most two weeks. Some features of Feature Driven Development methodology are the concern for quality, thanks to its monitoring. On the other hand also counteracts the failure situations or lack of program elements desired. This methodology proposes stages closing every two weeks and tangible results obtained to assess the project progress. Although, FDD does more emphasis on the design and construction phases than the requirements phase. The process (Fig. 3) for developing the application will consist of:

1. Development of a general model
2. Building a list of features
3. Planning by features
4. Design and construction of each feature

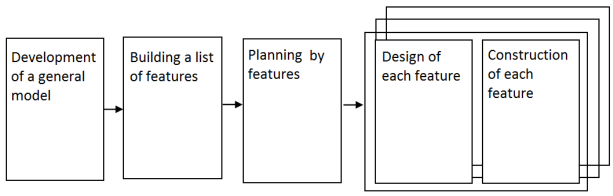


Fig. 3: Feature Driven Development.

5.2 Software Development Tool

For the treatment of images I am able to use many programming languages such as Python, Matlab, Java, C#, C++, among others. From all this I decided to do this project with Matlab for the facility that provides when processing images. In addition, Matlab works with images as arrays. Matlab has different features and functions: find similarities between images, binarization, projections, labeling, etc.

5.3 Databases

We have two databases of isolated musical symbols, facilitated by Dr. Ana Rebelo [13] and CVC [26], which will help me to classify the different symbols detected in my scores. The techniques used will be settled, finally deciding which gives the better results.

5.4 Pipeline overview

Figure 4 shows the pipeline stages: When a handwritten music score enters into the system, first of all it is preprocessed (Section 7). Next, each symbol will be classified (Section 8) into isolated (Section 9) or compound (Section 10) and recognized with the corresponding method.

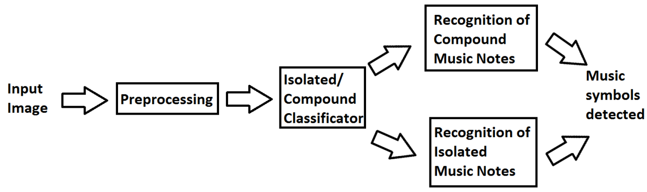


Fig. 4: Pipeline.

6 Temporal Planning

The whole project is divided in different tasks:

1. Planning: To build the list of features, choosing the programming language, preparing the development environment, and so on.

2. State of Art: To read and study articles from the literature to learn and understand the state of the art on recognition of music scores.

3. Hypothesis: First state the hypothesis to decide how to face the problem and then think possible solutions

4. Development and evaluation: Implement the solutions designed in the previous stage.

5. Documentation: Prepare all the documentation. It is composed by ICFHR paper, the previous and final report and the presentation.

Fig. 5: Temporal planning.

We have followed the initial planning. Although the initial planning was stipulated for 300 hours, later extra effort was added with the subject "work placement". I followed the same planning but adding more hours to each task. The total project has lasted approximately 500 hours.

7 Preprocessing

In the preprocessing, the system removes music braces and ties. In this step the system assumes that the input image is binary and the staff lines are already removed by using any of the staff removal methods in the literature [27].

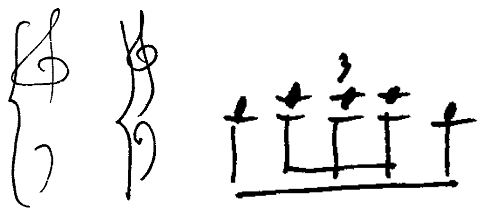
7.1 Brace and Tie removal

7.1.1 Brace removal

In polyphonic music scores, braces indicate that several staffs are played at the same time, such as in music scores for different instruments. Given that braces appear at the beginning, the system analyses the connected components at the beginning of the staffs. Following the musical notation theory, a brace must cross consecutive staffs. Thus, these elements are approximated to a straight line, and if the estimated line crosses several staffs, it is classified as a brace. Afterwards, braces are removed using the straight line estimation in order to avoid the deletion of other elements such as clefs. The removal of braces will ease the posterior recognition of music symbols, such as clefs and key-signatures. Figure 6a) shows some examples of braces where some of them are overlapping the clefs. For more details, see [28].

7.1.2 Tie removal

Long ties are used for adding expressivity in music performance. However, they can be easily misclassified as beams due to the handwriting style of the musician. Figure 6b) shows a problematic case, where the beam is disconnected from the stems. Therefore, I propose to detect and remove the long ties by analyzing the aspect ratio of the horizontally long connected components.



(a) (b)

Fig. 6: a) Examples of braces that are gathered with clefs. b) Beam easily confused as a tie. Image extracted from [28]

7.2 Segmentation

As soon as ties and braces have been removed, it is the moment to find each music symbol. First of all, with the image of the staff the system does do a horizontal projection to obtain a template (see Fig.7 b) with the regions for each staff. If some symbol is between two regions, the system look for connected components and take the region which has bigger area. Secondly, based on the results of the Regions Labeled, a vertical projection is applied (see Fig.7 c) to find each symbol (see Fig.7 d).

8 Isolated and Compound Music Notes Classification

Each symbol will be classified as Isolated Music Note or Compound Music Note depending on the number of detected note-heads.

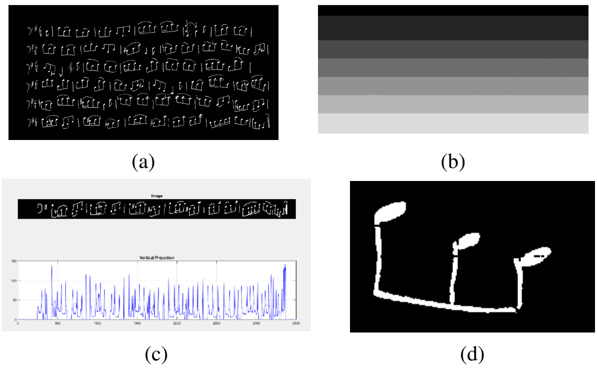


Fig. 7: a) Input image. b) Labeled regions. c) Vertical Projection. d) Segmented symbol.

8.1 Detection of Note-head

Note-heads play a key-role in music notes, since they provide the melody. Moreover, its the only common component in all type of music notes. Hence, detecting correctly a note-head is of key importance for the correct symbol construction. Figure 8 shows in red the different types of note heads that must be detected.

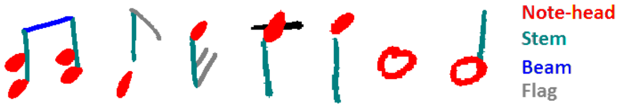


Fig. 8: Graphics Primitives.

Filled-in note-heads are detected using mathematical morphology. First, two elliptic structural elements are defined using different angles. Then a morphological closing is performed using both structural elements. Finally, blobs closer to a vertical line are considered filled-in note-heads. For the detection of white note-heads, the filled-in note-heads are first removed from the image. Then, the holes are filled so that white note-heads can be found using the same strategy. In both cases, too large blobs are rejected.

9 Recognition of Isolated Music Notes

If [0,1] Note-head is detected, this symbol is a candidate to be a simple Music symbol.

Fig. 9: Some Isolated Music Notes.

Each candidate will be recognized by a symbol descriptor (Zoning /BSM). Next to K-means extract the centroids of representative samples to streamline the process of K-NN.

9.1 Zoning

Zoning [29] computes the percentage of black pixels in each zone: an m x n grid is superimposed on the symbol image, and for each of the nxm zones, the average gray level is computed, giving a feature vector of length nxm. This method is fast and very useful in the recognition of printed symbols.

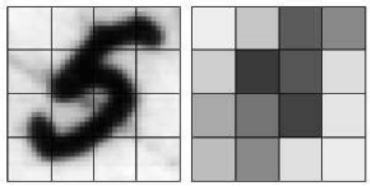


Fig. 10: Zoning. Image extracted from [29]

9.2 BSM

The second method is the Blurred Shape Model (BSM) [30] descriptor. BSM encodes the spatial probability of the shape. Firstly, the image is divided in a grid of n x n subregions, and each bin receives votes from the shape points in it and also from the shape points in the neighboring bins. The output descriptor is a vector histogram where each position corresponds to the amount of shape points in the context of the sub-region. The resulting vector histogram, obtained by processing all feature points, can be normalized in the range [0..1] to obtain the probability density function of n x n bins.For further details, see [12] .

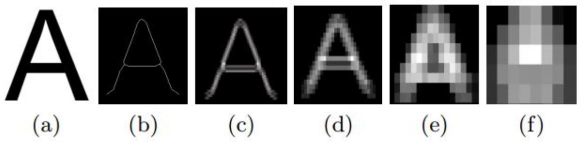


Fig. 11: (a) Input image. (b) Thinned image. (c) 64 regions blurred shape. (d) 32 regions blurred shape. (e) 16 regions blurred shape. (f) 8 regions blurred shape. Image extracted from [31]

9.3 K-means

The algorithm K-means clustering is a method that aims to partition a set of n observations into k clusters in which each observation belongs to the cluster with the nearest mean.

9.4 K-NN

The k-NN (or nearest neighbor) is a non-parametric method used for classification, the output is a class membership. The image is classified by a majority vote of its neighbors, with the object being assigned to the most common class among its k nearest neighbors.

10 Recognition of Compound Music Notes

If [2,∞) Note-head are been detected, this symbol is candidate to be a Compound Music Note. Once I know that is a compound music note the following step its detect the rest of graphic primitives.

10.1 Detection of Graphics Primitives

The starting point to construct the proposed hierarchical representation is the detection of basic primitives that defines the musical vocabulary or compound notes. These basic primitives are created by means of simple detectors.

10.1.1 Vertical lines detection

Vertical lines are key elements that are mainly used to represent Stems and Bar lines. Since music notes are mainly composed by note-heads, stems, beams and flags (see Fig.8), I must identify the bar lines so that I can keep the rest of vertical lines as stem candidates. For this task, I first detect all the vertical lines using a median filter, and then, I analyze them to identify the bar lines. The bar line identification consists of two steps.

- *Properties checking*: The vertical line is kept as a bar line candidate if it (almost) crosses all the staff and it has no blobs (note-heads) at its extrema points.

- *Consistency checking*: The bar lines in the same page must have similar length and must cross the same staffs. Therefore, the consistency is analyzed as follows. First, I vertically sort the bar line candidates using their centroid. Here, one candidate is an outlier if its length is very different from the candidates in the same line. These outliers are analyzed just in case they have not been correctly detected so they must be joined with other vertical lines. Otherwise, they are rejected.

10.1.2 Beam detection

The beam's appearance highly depends on the melody. Consequently, a descriptor based on densities, profiles or gradients will be unstable. For this reason, I propose the detection of beams by adapting a pseudo-structural descriptor [32] for handwritten word spotting. The feature vector is created from the information from every key-point in the word. For each key-point, the characteristic Loci Features encode the frequency of intersection counts following a certain direction path. Thus, the shape of the strokes is not taken into account.

For the detection of beams, the idea is to modify the pseudo-structural descriptor as follows. For each pair of consecutive detected note-heads (and stems), I take the region in between, and divide it into 2 parts (left and right). Then, I compute the characteristic Loci Features in the vertical direction (i.e. the number of transitions). Finally, I take the statistical mode (the most frequent value), which indicates the amount of beams that link each pair of notes. In this way, the descriptor is invariant to the beam appearance and orientation.

10.2 Perceptual Grouping

Once the graphics primitives are detected, the next step consists in grouping them to recognize the compound music notes. For this purpose, I first create a hierarchical representation of primitives, and then I validate the different grouping hypothesis using syntactical rules.

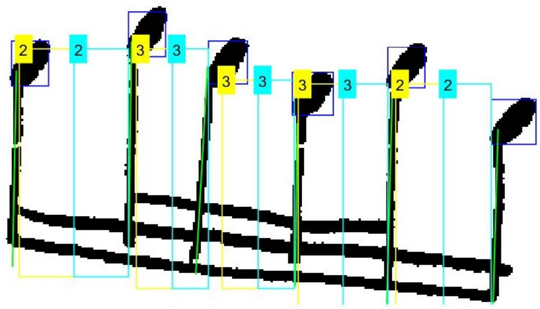


Fig. 12: Cropped Image with primitives detected.

10.2.1 Hierarchical representation

Inspired in the perceptual grouping techniques for text detection [20] and object recognition [21], I propose to built a dendrogram to hierarchically represent the graphics primitives (see Fig. 13). In my case the criteria for grouping is the proximity of the graphics primitives, which means that the coordinates of the primitives' centers are used as features to create the hierarchical clustering.

Since a music note must contain at least one note-head, I use the detected note-head candidates as seeds to start the grouping in a bottom-up manner. Thus, I can easily avoid the creation of many non-meaningful grouping regions.

Notice that the different grouping hypothesis can overlap. For instance, a chord is composed of several note-heads that share the same stem (e.g. see the first note in Fig. 8). Consequently, this stem belongs to more than one group hypothesis.

10.2.2 Validation of grouping hypothesis

Once I have the different hypothesis of groups of primitives, the next step is to validate them. In case of text detection, the grouping validation could be performed by trying to recognize the text. For example, a grouping hypothesis could be accepted whenever an OCR can recognize the word.

However, in this case, trying to recognize the compound notes as a whole is not possible because the creation of a dictionary of music notes is unfeasible: there are almost-infinite combinations of compound notes. Moreover, even if it was possible, I would need an huge amount of samples to train a shape recognizer. Therefore, I propose to validate each one of the grouping hypothesis through the following music notation rules:

- Whole note = {[white-note-head]+ }.

- Half note = {[white-note-head]+, stem}.

- Quarter note = {[filled-in-note-head]+, stem}.

- 8th note = {[filled-in-note-head]+, stem, beam}.

- 16th note = {[filled-in-note-head]+, stem, beam, beam}.

- 32th note = {[filled-in-note-head]+, stem, beam, beam, beam}.
- 64th note = {[filled-in-note-head]+, stem, beam, beam, beam, beam}.

The symbol + indicates that minimum one appearance of this primitive is required. In summary, only the grouping hypothesis that can be validated using these rules will remain. All the other hypothesis will be rejected. See an example in Fig 13.

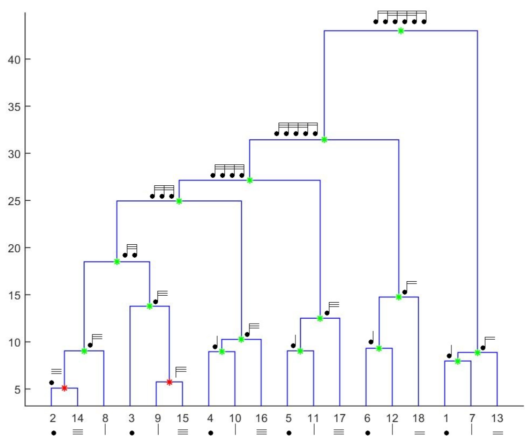


Fig. 13: Validation hypothesis (Dendrogram) of Fig. 12 .

11 Evaluation

11.1 Isolated and Compound Music Notes Classification

For the experiments, I have selected a subset of the CVC-MUSCIMA dataset [33].

Table 1 shows the experimental results. The first column indicates the music page that has been used (e.g. 'w5-p02' means page 2 from writer 5). The second column shows the detection of isolated music notes, and the third column shows the detection of compound music notes. The metric computed is the classification rate (number of correctly detected elements divided by the number of detected elements).

11.2 Isolated Music Notes

For the experiments of Isolated Music Notes, I have selected a test database with 30 deferents symbols (quarter-notes, half-notes, clefs, rests...). For the training I used a symbol database of the CVC [26] and a symbol database of Ana Rebelo [13]. Each image has been tested with diferents modifications (applying thinning, with borders, synthetic and handwritten images separately, among others). BSM and Zoning have been tested with different parameters from n=5 to n=10 with the centroids returned by K-means, from K=5 to K=10.

Table 2 shows the experimental results. The first column indicates the modifications applied to the images. The second column indicates the best configuration of Zoning and the classification rate. And the last column indicates the best configuration of BSM and the classification rate.

TABLE 1: Results. The classification between Isolated Music Notes and Compound Music Notes are shown in terms of classification rate (P). All results are between [0-1].

Score	P. Isolated	P. Compound	Total
w5-002	0,97	0,9	0,96
w5-010	0,99	0,78	0,9
w5-011	1	0,9	0,96
w5-012	0,99	0,92	0,97
w10-002	1	0,66	0,96
w10-010	1	0,71	0,87
w10-011	0,95	0,95	0,95
w10-012	1	0,44	0,87
w17-012	1	0,72	0,91
w38-012	1	0,82	0,95
Mean	0,98	0,77	0,92

TABLE 2: Results. The detection of Isolated Music Notes in terms of best configuration by BSM and Zoning and the classification rate. All results are between [0-1].

Configuration	Zoning	BSM
Without changes	Zoning8-K10 0.43	BSM10-K6 0,53
Borders	Zoning6-K10 0.31	BSM9-K8 0,45
Thining	Zoning9-K5 0.46	BSM10-K6 0,50
Syn. symbol as a centroide	Zoning9-K5 0.46	BSM10-K6 0,50
Syn. and hw. symbols separatly - Without changes	Zoning9-K10 0.70	BSM7-K9 0,70
Syn. and hw. symbols separatly - Thining	Zoning8-K7 0.50	BSM9-K7 0,53

From the results I observe that the detection of music notes obtains, always, a better results with BSM than zoning. For this reason, the finals experiments will be executed with BSM7-K9.

11.3 Compound Music Notes

For the experiments, I use the same subset than in Section 11.1. Concretely, I have manually created the ground-truth of 10 music pages, which contain a total of 1932 music notes. The music scores are from 4 different writers, mostly polyphonic music (containing several voices and chords).

Table 3 shows the experimental results. The first column indicates the music page that has been used (e.g. 'w5-p02' means page 2 from writer 5). The second column indicates whether the score is polyphonic or monophonic. The third and forth columns show the detection of note-heads, whereas the last two columns show the detection of music notes (e.g. 8th note, 16th note, etc.). The metrics computed are the Precision (number of correctly detected elements divided by the number of detected elements), and Recall (number of correctly detected elements divided by the number of elements in the dataset).

TABLE 3: RESULTS. THE DETECTION OF NOTE-HEADS AND MUSIC NOTES ARE SHOWN IN TERMS OF PRECISION (P) AND RECALL (R). ALL RESULTS ARE BETWEEN [0-1].

Score	Polyphonic	Note-heads		Notes	
		P	R	P	R
w5-002	No	0,73	0,56	0,54	0,41
w5-010	Yes	0,69	0,59	0,42	0,36
w5-011	No	0,68	0,68	0,52	0,52
w5-012	Yes	0,71	0,65	0,53	0,48
w10-002	No	0,68	0,56	0,35	0,29
w10-010	Yes	0,66	0,62	0,45	0,42
w10-011	No	0,72	0,57	0,48	0,38
w10-012	Yes	0,64	0,58	0,46	0,42
w17-012	Yes	0,66	0,61	0,56	0,52
w38-012	Yes	0,67	0,70	0,60	0,62
Mean	-	0,68	0,61	0,49	0,44

From the results I observe that the detection of music notes obtains a mean Precision and Recall in most cases is close to 50%. The main reason is that the detection of note-heads (which are used as seeds in the grouping) is sensitive to the handwriting style. For example, the head-note detector misses almost 40% of the note-heads. Consequently, the detection of music notes will be always lower that this value.

Our method has been compared with PhotoScore [6], a commercial OMR software able to recognize handwritten music scores. Figures 15 and 16, shown in Appendix A.1, show qualitative results from both approaches. As it can be noticed, PhotoScore performs very well in easy parts, whereas its performance decreases considerably in case of complex compound music notes. In this aspect, my approach is much more stable.

Table 4 shows quantitative results of my method compared to the ones obtained from PhotoScore. As it can be seen, my method approximates the recognition of compound music notes 'w38-012'. Contrary, the differences in the recognition of the 'w10-010' score are very high. This is mainly due to the different accidentals found in the compound music symbols such as sharps or naturals which creates confusion in the dendrogram. See an example on figure 14.

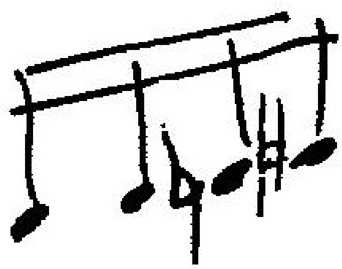


Fig. 14: Compound Music Note with accidentals (Natural & Sharp)

TABLE 4: COMPARISON WITH THE COMMERCIAL PHOTOSCORE OMR SOFTWARE. DETECTION OF MUSIC NOTES IN TERMS OF PRECISION (P) AND RECALL (R). ALL RESULTS ARE BETWEEN [0-1].

Score	PhotoScore		Our method	
	P	***R***	***P***	***R***
w10-010	0,63	0,61	0,45	0,42
w38-012	0,69	0,74	0,6	0,62

TABLE 5: RESULTS. THE DETECTION OF ISOLATED AND COMPOUND MUSIC NOTES ARE SHOWN IN TERMS OF PRECISION (P) AND RECALL (R). ALL RESULTS ARE BETWEEN [0-1].

Score	Polyphonic	Isolated		Compound		Total	
		P	***R***	***P***	***R***	***P***	***R***
w5-002	No	0,46	0,62	0,54	0,41	0,49	0,52
w5-010	Yes	0,63	0,65	0,42	0,36	0,47	0,42
w5-011	No	0,67	0,42	0,52	0,52	0,55	0,49
w5-012	Yes	0,60	0,49	0,53	0,48	0,55	0,49
w10-002	No	0,46	0,69	0,35	0,29	0,51	0,46
w10-010	Yes	0,54	0,55	0,45	0,42	0,47	0,44
w10-011	No	0,56	0,40	0,48	0,38	0,49	0,38
w10-012	Yes	0,49	0,39	0,46	0,42	0,47	0,40
w17-012	Yes	0,44	0,55	0,56	0,52	0,52	0,53
w38-012	Yes	0,65	0,56	0,60	0,62	0,62	0,60
Mean	-	0,55	0,53	0,49	0,44	0,51	0,47

In any case, it must be said that this comparison is not completely fair. PhotoScore has some features to improve its performance that are not considered in my method. First, PhotoScore is a complete OMR system that recognizes the whole score, which probably uses training data to deal with the variability in the handwriting style. Since it recognizes all music symbols (including clefs, accidentals and rests), it can use syntactic rules for validation. For instance, the system can recognize the time signature and then validate the amount of music notes at each bar unit (which is used to solve ambiguities).

11.4 Final Recognition

Table 5 shows the experimental results. The third and forth columns show the detection of Isolated Music Notes (e.g. quarter, half, whole, etc.), the fifth and sixth columns, show the detection of Compound Music Notes (e.g. 8th note, etc.), whereas the last two columns show the total results. The metrics computed are the Precision and Recall.

From the results I observe that the detection of Isolated music notes obtains a mean Precision and Recall below 60%, when in the isolated experiments I had 70% of Precision. The main reason is that the Isolated experiments images were images only one music symbol, because the cropping was manually done. Actually, the cropping is automatic and in a subimage it may appear more than one symbol, such as a treble clef and two sharps. For this reason I do not obtain the 70% of Precision and Recall.

12 Conclusion

After the development of the project and analyze the results, I am able to say that the goals and subgoals have been achieved.

In this work I have proposed a method for recognizing music notes in handwritten music scores. My method is composed of an isolated music notes recognition and a compound music notes recognition with hierarchical representation of graphics primitives, perceptual grouping rules and a validation strategy based on music notation. Since my compound recognizer does not use any training data, the experimental results are encouraging, especially when compared with a commercial OMR software.

As a future work, firstly, I would like to improve the detection of note-heads, since it is clearly limiting the performance of my method. In this sense, a more sophisticated key-point detector for note-heads should be investigated. Moreover, I also plan to incorporate syntactical rules (e.g. time measure checking). Finally, I would make an user interface to be able to convert the music score to a digital format. This tool will be semiautomatic because if I do not get a certain percentage of confidence in classification of a symbol then, the user will fill that part.

Acknowledgments

Firstly, I would like to express my sincere gratitude to my advisor Dr. Alicia Fornés for the continuous support of my Bachelor's Degree Final Project, for her patience, motivation, and for introducing me to the research world. Her guidance helped me in all the time of research and writing of this paper. I could not have imagined having a better advisor and mentor for my Bachelor's Degree Final Project.

My sincere thanks also goes to Pau Riba, without his precious support it would not be possible to conduct this research. Additionally, I would like to thank Lluis Gomez for the several suggestions on the hierarchical representation.

Last but not the least, I would like to thank my family, friends and classmates for supporting me spiritually throughout writing this paper and my life in general.

References

[1] D. Bainbridge and T. Bell, "The challenge of optical music recognition," *Computers and the Humanities*, vol. 35, no. 2, pp. 95–121, 2001.

[2] A. Rebelo, I. Fujinaga, F. Paszkiewicz, A. Marcal, C. Guedes, and J. Cardoso, "Optical music recognition: state-of-the-art and open issues," *International Journal of Multimedia Information Retrieval*, vol. 1, no. 3, pp. 173–190, 2012.

[3] A. Fornés and G. Sánchez, "Analysis and recognition of music scores," in *Handbook of Document Image Processing and Recognition*. Springer-Verlag London, 2014, pp. 749–774.

[4] D. Pruslin, "Automatic recognition of sheet music, phd thesis," Massachusetts, USA, 1966.

[5] D. Prerau, "Computer pattern recognition of standard engraved music notation, phd thesis," Massachusetts, USA, 1970.

[6] "Photoscore." [Online]. Available: http://www.neuratron.com/photoscore.htm

[7] "Sharpeye." [Online]. Available: http://www.visiv.co.uk/

[8] A. Rebelo, F. Paszkiewicz, C. Guedes, A. R. S. Marcal, and J. S. Cardoso, "A method for music symbols extraction based on musical rules," in *Proceedings of BRIDGES*, 2011, pp. 81–88.

[9] J. C. Pinto, P. Vieira, and J. M. Sousa, "A new graph-like classification method applied to ancient handwritten musical symbols," *International Journal on Document Analysis and Recognition*, vol. 6, no. 1, pp. 10–22, 2003.

[10] L. Pugin, "Optical music recognition of early typographic prints using hidden markov models," in *International Conference on Music Information Retrieval*, 2006, pp. 53–56.

[11] A. Fornés, J. Lladós, G. Sánchez, and D. Karatzas, "Rotation invariant hand drawn symbol recognition based on a dynamic time warping model," *International Journal on Document Analysis and Recognition*, vol. 13, no. 3, pp. 229–241, 2010.

[12] S. Escalera, A. Fornés, O. Pujol, P. Radeva, G. Sánchez, and J. Lladós, "Blurred Shape Model for binary and grey-level symbol recognition," *Pattern Recognition Letters*, vol. 30, no. 15, pp. 1424–1433, 2009.

[13] A. Rebelo, G. Capela, and J. S. Cardoso, "Optical recognition of music symbols: A comparative study," *International Journal on Document Analysis and Recognition*, vol. 13, no. 1, pp. 19–31, 2010.

[14] K. Ng, "Music manuscript tracing," in *Graphics Recognition Algorithms and Applications*. Springer, 2001, pp. 330–342.

[15] H. Miyao and M. Maruyama, "An online handwritten music symbol recognition system," *International Journal of Document Analysis and Recognition*, vol. 9, no. 1, pp. 49–58, 2007.

[16] J. Calvo-Zaragoza, J.; Oncina, "Recognition of pen-based music notation with probabilistic machines," in *Proceedings of the 7th International Workshop on Machine Learning and Music*, Barcelona, Spain, 2014.

[17] "Staffpad." [Online]. Available: http://www.staffpad.net/

[18] "Myscript music." [Online]. Available: http://myscript.com/technology/

[19] "Notateme." [Online]. Available: http://www.neuratron.com/notateme.html

[20] L. Gomez and D. Karatzas, "Multi-script text extraction from natural scenes," in *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*. IEEE, 2013, pp. 467–471.

[21] N. Ahuja and S. Todorovic, "From region based image representation to object discovery and recognition," in *Structural, Syntactic, and Statistical Pattern Recognition*. Springer, 2010, pp. 1–19.

[22] A. Fornés, J. Lladós, G. Sánchez, and D. Karatzas, "Rotation invariant hand-drawn symbol recognition based on a dynamic time warping model", journal="international journal on document analysis and recognition (ijdar)," vol. 13, no. 3, pp. 229–241, 2010.

[23] H. Kato and S. Inokuchi, *Structured Document Image Analysis*. Springer-Verlag, 1991, ch. A recognition system for printed piano music using musical knowledge and constraints, pp. 435–455.

[24] S. Macé, É. Anquetil, and B. Coüasnon, "A generic method to design pen-based systems for structured document composition: Development of a musical score editor," in *Proc. of the 1st Workshop on Improving and Assessing Pen-Based Input Techniques*, 2005, pp. 15–22.

[25] A. Baró, P. Riba, and A. Fornés, "Towards the recognition of compound music notes in handwritten music scores," in *International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 2016, submitted.

[26] A. Fornés, J. Lladós, and G. Sánchez, "Old handwritten musical symbol classification by a dynamic time warping based method," in *GREC*, 2007.

[27] M. Visani, V. C. Kieu, A. Fornés, and N. Journet, "Icdar 2013 music scores competition: Staff removal," in *12th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2013, pp. 1407–1411.

[28] P. Riba, A. Fornés, and J. Lladós, "Towards the alignment of handwritten music scores," in *Eleventh IAPR International Workshop on Graphics Recognition (GREC)*, 2015.

[29] A. Fornés, "Analysis of old handwritten musical scores," pp. 67–68, 2005.

[30] A. Fornes and J. Llados, "A symbol-dependent writer identification approach in old handwritten music scores," pp. 634–639, Nov 2010.

[31] A. Fornés, S. Escalera, J. LLadós, G. Sánchez, P. Radeva, and O. Pujol, "Handwritten symbol recognition by a boosted blurred shape model with error correction," pp. 13–21, 2007.

[32] J. Lladós, M. Rusinol, A. Fornés, D. Fernández, and A. Dutta, "On the influence of word representations for handwritten word spotting in historical documents," *International journal of pattern recognition and artificial intelligence*, vol. 26, no. 05, p. 1263002, 2012.

[33] A. Fornés, A. Dutta, A. Gordo, and J. Lladós, "Cvc-muscima: a ground truth of handwritten music score images for writer identification and staff removal," *International Journal on Document Analysis and Recognition*, vol. 15, no. 3, pp. 243–251, 2012.

APPENDIX

A.1 Qualitative Results

Figures 15 and 16, show qualitative results from two great stave of different music scores.

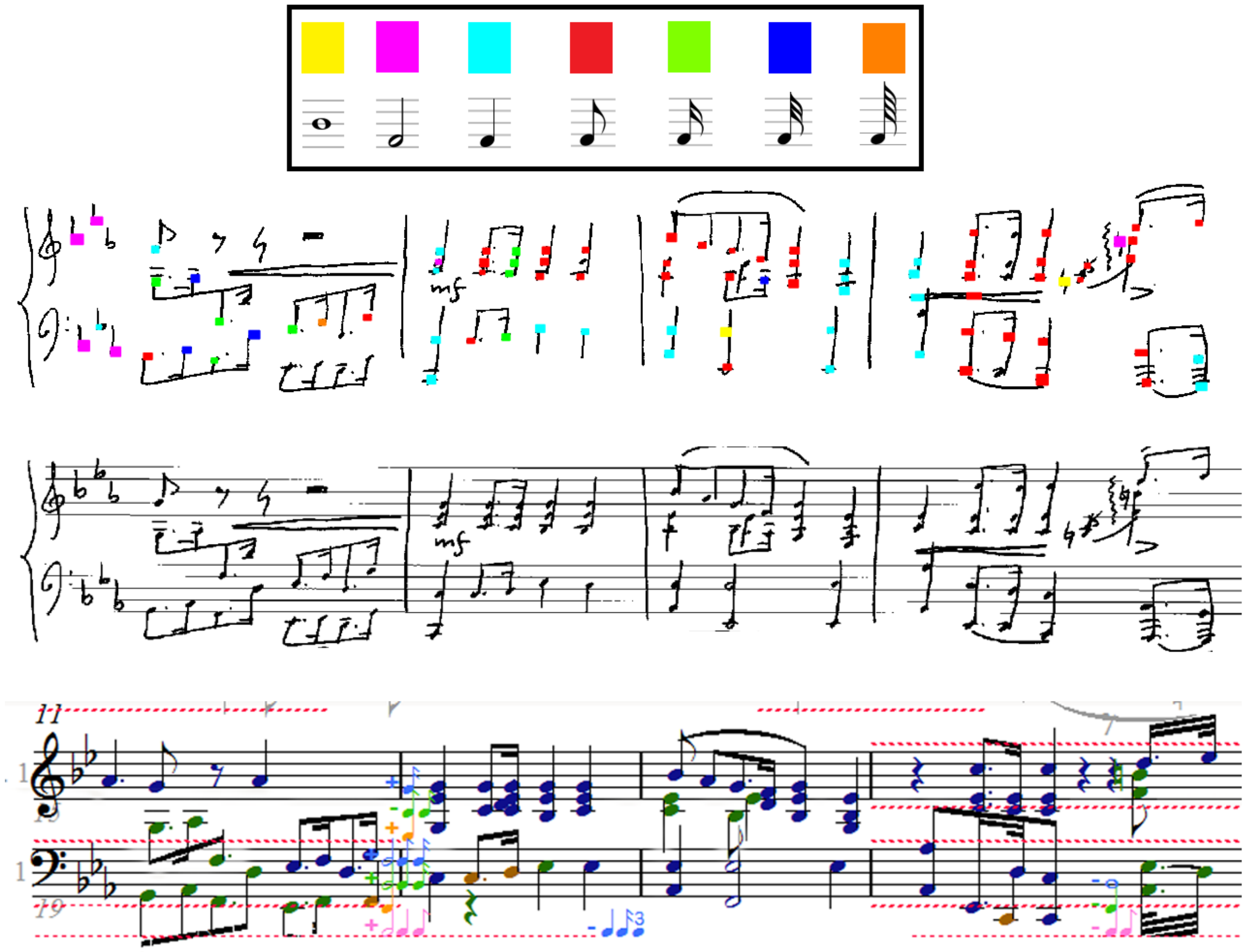


Fig. 15: Results on ‘w10-p10’. First row: my method. Second row: original image. Third row: PhotoScore results.

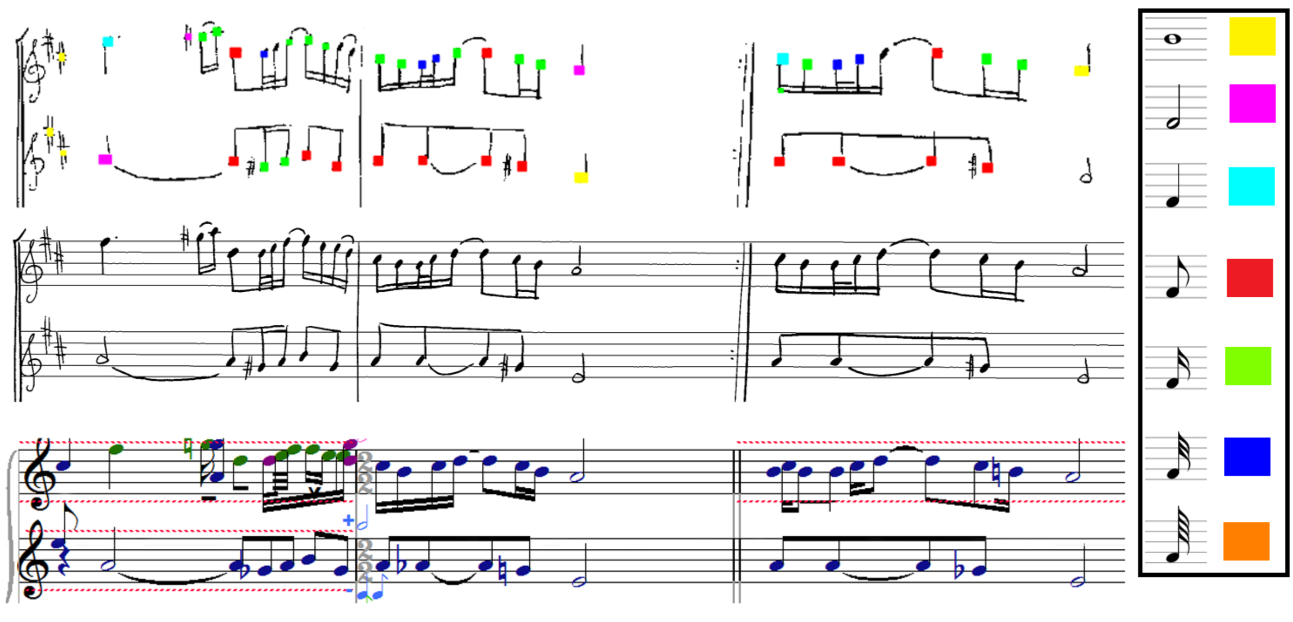


Fig. 16: Results on‘w38-p012’. First row: my method. Second row: original image. Third row: PhotoScore results.