Symbol Classification Approach for OMR of Square Notation Manuscripts

Carolina Ramirez
Waseda University
ramirez@akane.waseda.jp

Jun Ohya
Waseda University
ohya@waseda.jp

ABSTRACT

Researchers in the field of OMR (*Optical Music Recognition*) have acknowledged that the automatic transcription of medieval musical manuscripts is still an open problem [2, 3], mainly due to lack of standards in notation and the physical quality of the documents. Nonetheless, the amount of medieval musical manuscripts is so vast that the consensus seems to be that OMR can be a vital tool to help in the preserving and sharing of this information in digital format.

In this paper we report our results on a preliminary approach to OMR of medieval plainchant manuscripts in square notation, at the symbol classification level, which produced good results in the recognition of eight basic symbols. Our preliminary approach consists of the preprocessing, segmentation, and classification stages.

1. INTRODUCTION

Several groups are currently working to build digital archives and catalogues using digital technologies [10, 11, 12, 13, 14], of the huge number of early musical manuscripts accessible from multiple sources. The lines of research of these groups in early music information retrieval range from the design of web protocols for digital representation of scanned early music sources to the automatic transcription of those sources through adaptive techniques [2, 5, 9, 10]. Given the physical and semantic characteristics of many of these documents (degradation, non-standard notation, etc.), great variability is introduced to the data, and the subsequent analysis can be a quite difficult and time consuming task, usually requiring advanced expert knowledge. So, until very recently, those mentioned efforts were restricted mostly to build text catalogues and repositories of scanned images.

In the case of standard modern music notation, OMR has achieved high levels of accuracy, and there are several OMR systems commercially available [1, 15]. In the case of early music manuscripts, attempts to achieve good OMR results become more challenging as our sources go back in time. Still, researchers have extended their work to early music manuscripts, and in the past years we have observed advances in renaissance printed music and handwritten music [4, 5, 17], but still little has been re-ported about experimental results with western plainchant medieval sources [2]. The work done by the NEUMES project [10], and most recently by Burgoyne et al. [3], are among the few experimental results with this particular type of source. In [2] the problem of non-standard notation is mentioned as the most critical issue for early manuscript OMR. For this reason, we start our research by restricting the manuscripts in square notation to belong to the XIV century and later, when square notation was already an established practice and basic symbols were more standardized than in previous neumatic alphabets [16].

In this paper we aim to successfully classify the eight basic characters of western square notation, see Figure 1, using relatively simple and widely known image processing and pattern recognition algorithms. If this proves successful, we believe that more complex models, context information, and adaptive techniques can be used in the future to minimize the errors at the classification stage, to extend the span of examples that can be analyzed, i.e. less standard documents, and to include a whole semantic analysis.
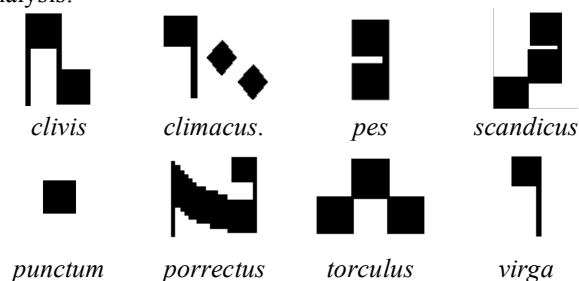

Figure 1: Square notation basic symbols.

Finally, it is necessary to mention that a big concern in this research area is the evaluation methods to be used. Symbol classification can be evaluated using the usual techniques, but creating a ground-truth for a full manuscript (where even the experts sometimes disagree) would require an effort that is beyond the scope of this paper.

2. OUTLINE

In section 3 we describe the preprocessing stage, which includes binarization of the manuscript image, location of staff lines and staves that define our ROI (*Region of Interest*), and stave deskewing. In section 4 we describe our segmentation and classification strategy. Lastly, in section 5 we present our conclusions and delineate some future work ideas.

3. PREPROCESSING

3.1 Binarization and ROI Extraction

As we said above, one of the biggest difficulties in analyzing early music manuscripts comes from the high variability on the image data introduced by the deteriorated state of the documents [9]. Besides dealing with a non-standard notation or non-standard scanning methods, the physical condition of some documents (high degradation, discoloration, missing parts, etc.) calls for an adequate amount of preprocessing. Some possibilities for the preprocessing stage include filtering, spatial transforms (Hough transform has been proposed to correct staff line positions [5]), and adaptive thresholding.

In order to binarize and extract the ROI we implement the adaptive approach proposed by Gatos et al. in [6]. The main advantage of this method is that it is able to deal with degradations due to shadows, non-uniform illumination, low contrast, smear, and strain. The disadvantage is that it is a parametric method, and in order to obtain good results some amount of parameter tuning is required [4]. The steps include an initial denoising using a 3x3 Wiener filter, a rough foreground estimation using Sauvola's Local Adaptive Threshold, a background estimation, and a final local thresholding using the distance between the Wiener filtered image and the background estimation. We did not implement the up-sampling stage in [6], because preliminary tests showed that it was not critical to detect our ROI.

The original image I, the filtered image I_w, the background image I_b, and the final binary image I_f are shown in Figure 2.

Original image I Filtered imaged I_w
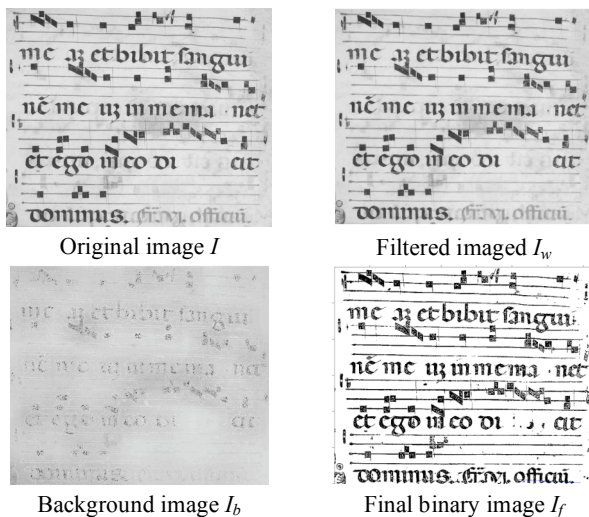
Background image I_b Final binary image I_f

Figure 2: Binarization stages.

We use the binary image I_f to detect our region of interest, the area of the image where the relevant symbols

are located, which in a musical document is a stave, i.e. a group of staff lines. There can be many staves in one document and we want to extract each one of them separately. This also helps to minimize the presence of text and drawings in the analyzed images, elements that could make our analysis more difficult.

As an initial approach, we perform a rough localization of the staff lines by first detecting the positions of all the lines in the document using polar Hough transform. After the lines are extracted, we use another feature to decide if a group of lines is a stave. This feature is the space between lines, which can be also estimated from the Hough transform. Here we use the hypothesis that spaces between staff lines on the same stave are relatively smaller than the space between staves. We use a k-means classifier to group the spaces and detect the staves. Figure 3 shows an example of stave detection. Only whole staves will be extracted, so staff lines that do not form a complete stave are not considered as part of the ROI.

Figure 3: Stave detection.

In Figure 3 it can be noticed that the whole length of the stave is not detected. To solve this problem we use heuristics based on the inter-staff line and inter-staves spaces and the dimensions of the image.

3.2 Staves Deskewing

Many OMR algorithms assume that staff lines are horizontal, but this is not necessarily true in old manuscripts.

Figure 4: Aligned staves.

In order to facilitate the analysis, and in case we want to apply standard OMR techniques, it is useful to horizontally align the images as much as possible. This can be done with the information already obtained from the Hough transform, by rotating against the Hough angle. The result of applying this rotation can be seen in Figure

4. Note that this approach does not address the issue of deformed staff lines.

4. SEGMENTATION AND CLASSIFICATION

As explained in the introduction, we aim to obtain good symbol classification results while at the same time using a relatively simple methodology. In general, the standard approach is to binarize the document and then segment and classify the symbols using binary representations. We cannot use this approach because, even though the binarization we used above allows us to find the region of interest in the image, it is not accurate enough to conserve all the pixel information of the symbols across all the documents in our database. Hence, we carry out the segmentation directly from the extracted staves in grayscale. Due to the difficulty in removing lines from heavily degraded and deformed documents, we decided to skip the staff lines removal stage, and thus avoid a pixel-wise approach for symbol segmentation. Instead, we detect and segment whole symbols using pattern matching via correlation, and then we use a SVM (*Support Vector Machine*) to classify the symbols from gradient-based features.

4.1 Segmentation

We use normalized correlations on each stave image to match an artificially generated binary pattern of each symbol to the regions where that symbol potentially appears. Some of the binary patterns can be seen in Figure 1, but the classes that present more variability in size and geometrical distribution (*pes*, *torculus*, *porrectus*, *clivis*) are also divided in subclasses. These patterns were applied in 3 different scales, based in the height of the stave, to each stave image. After this process, a set of detected candidates is obtained. These candidates are the input for the SVM. An example of this process is shown in Figures 5, 6, and 7.
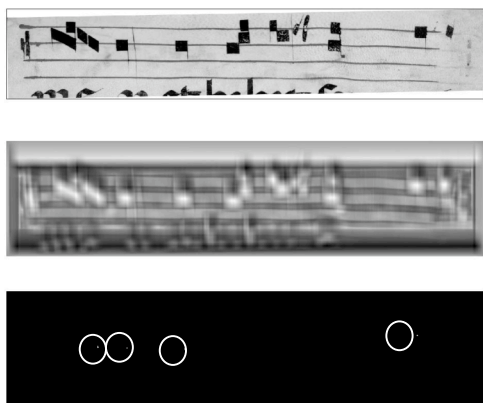

Figure 5: From top to bottom. Grayscale stave, normalized correlation image, and peaks of the correlation image.
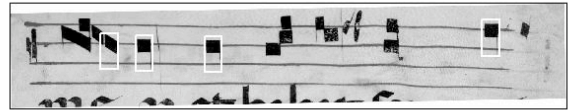

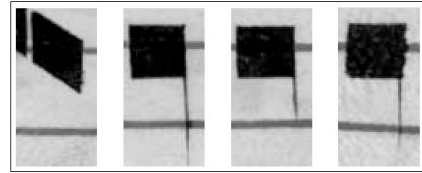**Figure 6**: Pattern Detection, class *virga*.

Figure 7: Segmented symbols, class *virga* (left, false detection).

After testing our detection algorithm in real documents [14], we observe that all basic symbols were detected with the binary patterns, but also many "false" candidates were extracted. These "false" candidates were mainly due to two causes: first, a basic pattern is actually part of another one, and second, a geometrical configuration similar to the basic pattern is formed by certain elements in the document. Examples of both conditions are shown in Figure 8.

Figure 8: Left, false *pes* detection (part of *scandicus*). Right, false *torculus* detection (part of *porrectus flexus*).

4.2 Classification

For Classification purposes, 1334 sample images of the 8 basic symbols were manually segmented and labeled from 47 sheets of music available at the Digital Scriptorium [14]. These sources are square notation manuscripts from the XIV to the XVII centuries (to avoid transitional times [16]), and from different geographical locations (Spain, Germany, Italy, etc.).

A size and position normalization using aspect ratio was performed on the samples [7], and 4 directional Sobel masks were applied to them (horizontal, vertical, left-diagonal, and right-diagonal) to obtain the gradient-based features used for classification. These Sobel images were divided in 96 blocks, and the mean gradient for each block was calculated. Finally, all the values were stacked in a feature vector [8].

We trained a SVM with a quadratic kernel function, and we tested it using cross-validation. The training was made using a *one-against-all* approach, thus obtaining a classifier for each of the eight classes. A simple voting algorithm is used to decide the final class from the outputs of the eight independent classifiers. Three experi-

ments were conducted, each with a different type of input. In the first experiment, we used grayscale samples without any quality enhancement, in the second experiment we used grayscale samples with contrast enhancement, and in the third experiment we used binary samples. Results are shown in Table 1.

Sample	Recall
Binary	0.8453
Grayscale	0.9208
Contrast enhanced	0.9610

Table 1: Classification rates for SVM cross-validation experiments. Values range from 0 to 1.

Table 2 shows the test results from 3000 independent examples, by class, for contrast-enhanced samples.

Class	Precision	Recall
Clivis	0.9331	0.9914
Climacus	0.9429	0.9519
Pes	0.9646	0.9542
Punctum	0.9674	0.9132
Porrectus	0.8476	0.8580
Scandicus	0.8667	0.8228
Torculus	0.9261	0.9482
Virga	0.9744	0.9311

Table 2: Classification results for contrast-enhanced samples. Values range from 0 to 1.

The candidates extracted from Section 4.1 were tested in the most successful of the three SVMs, with good classification rates. In the case of "false" candidates, the classifier is currently not capable of discern them as a different class, i.e. a class of "wrong" samples independent of the 8 basic classes.

5. SUMMARY AND DISCUSSION

We believe that our results, while not being completely conclusive, show that using a gradient-based feature generates good classification results of square notation at the symbol level provided the results from both detection and segmentation stages are good. When combining the detection stage with the classification stage, the performance is degraded by the presence of "false" detections obtained with the normalized correlation pattern matching. However, even if these results are not ideal, we consider that the errors in the classification of the "false" candidates can be reduced if we introduce two valuable elements into the analysis. The first element is the use of the redundancy in the detection, i.e. when two or more candidates are extracted from similar or overlapping positions in the image; the second element is the use of the context in which the symbol is found. In the first case, the sole presence of redundancy will alert us to the occur-

rence of an abnormal situation, and therefore allow us to act on it accordingly. In the second case, context information can be used to minimize errors: think of a basic pattern being part of another (for the worst case scenario, think of a *punctum*!). In that case, observing the context is essential to obtain complete information about the symbol under analysis, and be able to determine its correct class.

In terms of future work, our first concern is to improve the segmentation via pattern matching, without renouncing to other segmentation techniques. It is quite intuitive to imagine that some classes are more difficult to deal with. For instance, we observed that in many cases the classes *virga* and *punctum* were detected as the other, which makes us think that the characteristic stem of the *virga* has a weak influence in the normalized correlation pattern matching.

Finally, we believe that a robust analysis of these manuscripts cannot be completely achieved without also taking in account semantic context information. In general terms, plainchant is a sequence of sounds and rhythmic patterns evolving in time, and as such, models or techniques that deal with time sequences look like an attractive alternative to complement the symbol-based analysis and improve error management strategies. We know that certain rules are observed in Gregorian Chant, so, if some probabilistic rules can be derived from its semantics, even soft ones, we would like to undertake that direction of research.

6. ACKNOWLEGMENTS

7. REFERENCES

[1] Bainbridge, D. and Bell, T. *The Challenge of Optical Music Recognition.* Computers and the Humanities, No 35, pp95-121. 2001.

[2] Barton, L.W. G., Caldwell, J. A. and Jeavons, P. G. *ELibrary of Medieval Chant Manuscript Transcriptions*. Proceedings of the 5yth ACM/IEEE Joint Conference on Digital Libraries (Digital Libraries Cyberinfraestructure for Research and Education). Association for Computing Machinery. 2005, pp320-329.

[3] Burgoyne, J.A., Y. Ouyang, T. Himmelman, J. Devaney, L. Pugin, and I. Fujinaga. *Lyric extraction and recognition on digital images of early music sources*. Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009) 2009.

[4] Burgoyne, J. A., L. Pugin, G. Eustace, and I. Fujinaga. 2007. A comparative survey of image binarisation algorithms for optical recognition on degraded musical sources. *Proceedings of International Conference on Music Information Retrieval*. Vienna. 509–12.

[5] Fornes, A., Llados, J. & Sanchez, G. *Primitive Se mentation in Old Handwritten Music Scores*. Lecture Notes in Computer Science, vol. 3926, pp. 279-290. 2006.

[6] Gatos, B., Pratikakis, I.E., Perantonis, S.J. *Adaptive degraded document image binarization*. Pattern Recognition, Vol.39, No. 3, pp. 317-327. March 2006.

[7] CL Liu, K Nakashima, H Sako, H Fujisawa. Handwritten Digit Recognition: Investigation of Normalization and Feature Extraction Techniques. Pattern Recognition, vol. 37, pp. 265-279.2004 .

[8] CL Liu, K Nakashima, H Sako, H Fujisawa. *Handwritten Digit Recognition: Benchmarking of State_of_the_Art Techniques*. Pattern Recognition, vol. 36, pp. 2271-2285.2003

[9] Pugin, L., Burgoyne, J.A. & Fujinaga, I. *MAP Adaptation to Improve Optical Music Recognition of Early Music Documents Using Hidden Markov Models*. Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007), pp. 513-16. Vienna, Austria.

[10] NEUMES Project
http://www.scribeserver.com/NEUMES/index.html

[11] CANTUS Database
http://publish.uwo.ca/~cantus/

[12] The CAO-ECE Project
http://www.zti.hu/earlymusic/cao-ece/cao-ece.html

[13] Cantus Planus Study group
http://www.cantusplanus.org/

[14] Digital Scriptorium
http://www.digital-scriptorium.org

[15] OMR Systems
http://www.informatics.indiana.edu/donbyrd/OMRSystemsTable.html

[16] Nota Quadrata
http://notaquadrata.ca/index.html

[17] Aruspix
http://www.aruspix.net/project.html

[18] Lewis E M 73:13v. Used by permission of the rare Book Department, Free Library of Philadelphia.