E-028

A System for the OMR of Ancient Musical Manuscripts

Carolina Ramirez† Jun Ohya†

1. Introduction

The Optical Music Recognition (OMR) of ancient musical manuscripts is still an open issue in the field of Musical Information Retrieval and Digital Image Processing. Features such as the lack of standards in notation and the physical condition of the manuscripts add to the difficulty in analyzing these documents.

In this paper we describe our proposed methodology for the OMR of a specific style of ancient musical symbols, known as *square notation*, and we present preliminary results of the symbol-based character recognition stage, for eight basic symbols.

2. Background

Several groups are currently working to build digital archives and catalogues using digital technologies [5, 6], 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 [1]. 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 [7, 8]. 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 [2], but still little has been reported about experimental results with western plainchant medieval sources. In [1] 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 [8].

Our final goal is to build an OMR system in 2 stages, as in Figure 1. The first stage is a symbol segmentation and classification block that uses whole grayscale symbols as samples in order to avoid an "expensive" pixel-level segmentation approach. The second stage is an error detection block based in probabilistic networks, whose inputs will be the outputs of the first stage.

† GITS, Waseda University

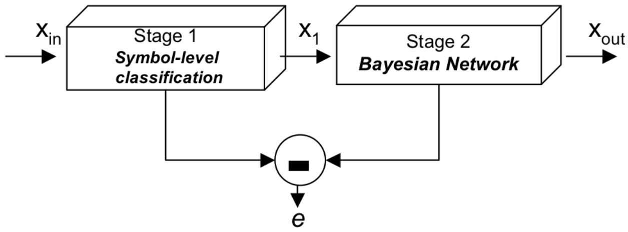

Figure 1: Two-stage general OMR system.

In this paper we present results for the first stage of our system. Our aim is to successfully classify the eight basic characters of western square notation, see Figure 2, using relatively simple and widely known image processing and pattern recognition algorithms.

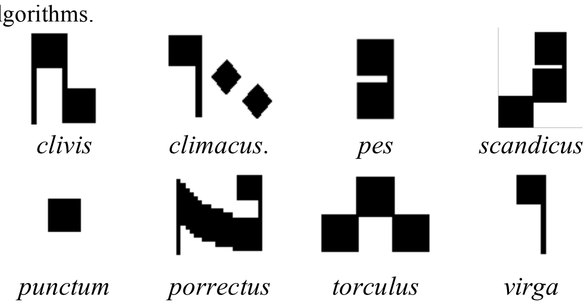

Figure 2: Square notation basic symbols.

3. Methodology

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. In order to binarize and extract the region of interest (ROI) we implement the adaptive approach proposed by Gatos et al. in [3]. 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 [2]. We use the binary image to detect our ROI, 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. We do this by using the Hough transform, and detecting groups of lines by a k-means clustering technique.

Finally, if necessary, we deskew the document based on a correction of the Hough angle of the staves.

3.2 Symbol 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 a 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.

We use normalized correlation on each stave image to match an artificially generated binary pattern, see Figure1, of each symbol to the regions where that symbol potentially appears. The classes that present more variability in size and geometrical distribution (*pes*, *torculus*, *porrectus*, *clivis*) are also divided in subclasses. The output of this process is a collection of symbol candidates for the SVM.

For classification purposes, over 3000 sample images of the 8 basic symbols were manually segmented and labeled from 47 sheets of music available at the Digital Scriptorium [6]. These sources are square notation manuscripts from the XIV to the XVII centuries. A size and position normalization using aspect ratio was performed on the samples [4], 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. 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

4. Results

Three experiments 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 3.2 were tested in the most successful of the three SVMs, with *good* classification rates. In 3.2 we also obtain "false" candidates, that is configurations of other symbols that resemble one of the basic symbols, but 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. Conclusion

The results of the symbol classification stage, even if still in need of refinement, are encouraging to our symbol-based OMR approach. At present, we are still developing our system, in particular, establishing a methodology to train a Bayesian network model for predicting the pitch and/or symbol occurrence of musical samples. This network uses a priori information given by certain stylistic restrictions specific to this kind of music. We believe that, by including this information in the system, some of the errors produced in the first stage can be detected, thus contributing to reduce the total error of the whole OMR system.

References

1. **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 Association for Computing Machinery. 2005, pp320-329.**
2. **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.**
3. **Gatos, B., Pratikakis, I.E., Perantonis, S.J. *Adaptive degraded document image binarization*. Pattern Recognition, Vol.39, No. 3, pp. 317-327. March 2006.**
4. **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 .**
5. **Cantus Planus Study group http://www.cantusplanus.org/**
6. **Digital Scriptorium http://www.digital-scriptorium.org**
7. **OMR Systems http://www.informatics.indiana.edu/donbyrd/OMRSystemsTable.html**
8. **Aruspix http://www.aruspix.net/project.html**