E-033

Proposed Method for OMR of Square Notation Manuscripts

Carolina Ramirez† Jun Ohya†

1. Introduction

The automatic transcription of medieval musical manuscripts is still an open problem in the field of Digital Image Processing [1]. Issues such as the poor physical state of the documents or the lack of a standard notation introduce great variability on the data to be analyzed.

In this paper we present a general framework for Optical Music Recognition (OMR) of medieval plainchant manuscripts in square notion.

Our method deals with three stages: staff lines detection, symbol classification and context-based error correction.

2. General OMR model

In [2] the authors propose a generic model, which simplifies an OMR system into four smaller tasks. The main advantage of this model is that allows comparing and contrasting different OMR systems, since most of them can be seen under this framework.

Figure 1 shows the OMR model introduced in [2].

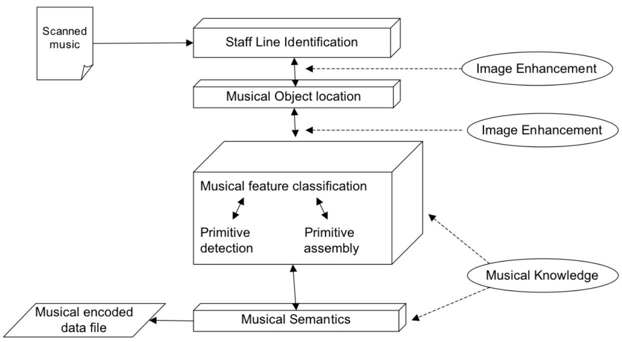


Fig. 1. Generic OMR model.

The four main tasks of the system are Staff line identification, Musical Object location, Musical feature classification and Musical Semantics; but we should mention that since ancient manuscripts are often in poor physical state, the image enhancement task at different stages is crucial to deal with the degradation of the documents. In this paper we present our strategy for binarization and staff lines detection.

3. Methodology

3.1 Preprocessing

The biggest difficulties of analyzing early music manuscripts come from the fact that the general conditions of the documents introduce high variability on the image data [3]. Besides dealing with a non-standard notation or non-standard scanning methods, the physical condition of some documents (high degradation, discoloration, missing parts, etc.) call for an adequate amount of preprocessing.

Some possibilities for the preprocessing stage include filtering, spatial transforms and adaptive thresholding

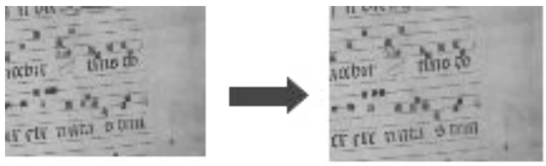
3.2 Binarization

For this stage we implement the adaptive approach proposed in [4]. The main advantages of this method are that requires a small amount of parameter tuning and is able to deal with degradations due to shadows, non-uniform illumination, low contrast, smear and strain. The steps are:

Filtering

An initial denoising step is made using a 3x3 Wiener filter.

Figure 2 shows the source image and the result of applying the Wiener filter.



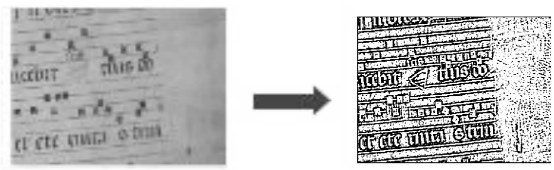
I_source I_wiener

Fig. 2. Initial denoising by Wiener filter.

Rough estimation of foreground

Sauvola's Local Adaptive Threshold is computed from the estimation of local means and standard deviations.

Here it is used to calculate a rough estimation of foreground pixels from the filtered image, as seen in Figure 3.



I_wiener I_sauvola

Fig. 3. Rough foreground estimation by Sauvola's Adaptive threshold.

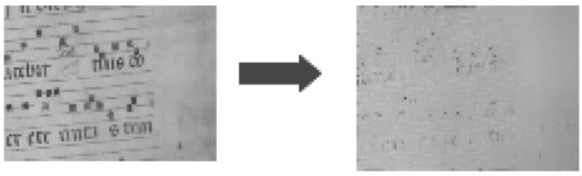
The advantage of using adaptive thresholding is that it improves the binarization results when the overall intensity values on the image are non-homogeneous and without strong illumination gradients. Sauvola's adaptive threshold is generally used with fixed parameters.

Background Estimation

In this step a map of background intensities is estimated from the filtered image and the foreground estimation from previous stages. Basically, if a pixel *(x, y)* is considered foreground in *I_sauvola*, then an estimated background value *B(x, y)* is calculated for that pixel via interpolation, according to values inside a neighborhood of *(x, y)*. Otherwise, the pixel is left as in

† GITS, Waseda University

the original image (i.e., is considered background). The result of this process can be seen in Figure 4.



I_wiener B(x,y)

Fig. 4: Original image and background estimation image B(x,y).

Final Thresholding

At this stage the distance *d* between *I_wiener* and *B(x,y)* is observed, and a threshold is calculated for each pixel according to *d*. In order to take in account areas of low contrast, a sigmoidal function is used to compute *d*.

This way:

$$T(x, y) = \begin{cases} 1 & \text{if } B(x, y) - I(x, y) > d(B(x, y)), \\ 0 & \text{otherwise.} \end{cases}$$

This is illustrated in Figure 5.

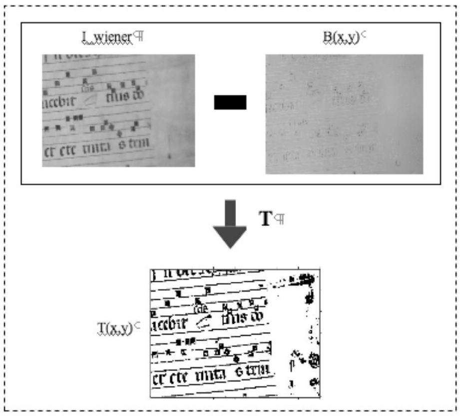


Fig. 5: Final thresholding stage. Binary image is calculated according to the distance between estimated background and original image.

There is room to improve the binary image obtained from this process, especially since default values were used, as proposed in [4]. An up-sampling and post-processing step is used in [4], but we have found that it might not be critical, either because the image already has good resolution and connectivity, or because the resultant binary image will be used only to find the location of staffs and symbols and not to extract features. Testing of this proposed stage is left as a future task.

3.3 Staff line detection

In a musical document staff lines are probably the main features and they can be easily observed. In our case, we first want to detect the region of interest where the symbols to be analyzed are located, which in a musical document is a stave, that is, staff lines grouped as a system. There can be many of these systems in one document.

As an initial approach, we perform a rough localization of the staves by first detecting lines using 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. Here we use the hypothesis that spaces between 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. The example in Figure 6 shows the detection of 2 staves of 5 staff lines each. The staff lines that do not form a system are not considered.

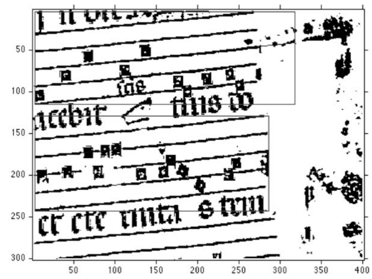


Fig. 6: Stave detection.

Many OMR algorithms assume that staff lines are horizontal, but as can be seen in Figure 6, in old manuscripts this assumption cannot be taken as granted. In order to apply some existent OMR techniques, is useful to deskew the images, which can be done with the information already obtained while applying the Hough transform. An example of the previous can be seen in Figure 7, where each stave has been processed separately.

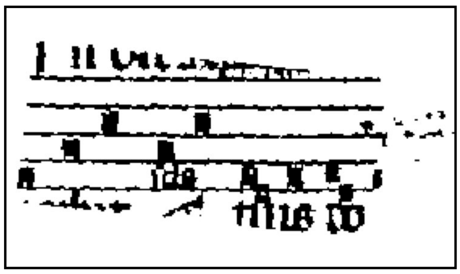


Fig. 7: Stave aligned horizontally.

4. Future Work

Future stages include musical object location, feature extraction and semantic interpretation. The step we propose as a novelty for this work is a filtering stage using probabilistic methods in order to reduce the errors of a standard OMR algorithm. The theoretical framework for this is still in a developing stage.

References

1. **Barton, L. W. G., Caldwell, J.A. & Jeavons, P. G. *E-Library of Medieval Chant Manuscript Transcriptions*. Proceedings of the 5th ACM / IEEE Joint Conference on Digital Libraries. Association for Computing Machinery. 2005. Pp. 320-329**
2. **Bainbridge, D. & Bell, T. *The Challenge of Optical Music Recognition.* Computers** and the Humanities, No 35, pp95-121. 2001.
3. **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.**
4. **Gatos, B., Pratikakis, I.E., Perantonis, S.J. *Adaptive degraded document image binarization*. Pattern Recognition, Vol.39, No. 3, pp. 317-327. March 2006.**