Old Handwritten Musical Symbol classification by a Dynamic Time Warping based method

Alicia Fornés, Josep Lladós, and Gemma Sánchez

Computer Vision Center, Dept. of Computer Science, Universitat Autònoma de Barcelona, 08193, Bellaterra, Spain
E-mail: {*afornes,sescalera,josep,gemma*}*@cvc.uab.es*

Abstract. A growing interest in the document analysis field is the recognition of old handwritten documents, towards the conversion into a readable format. The difficulties when working with old documents are increased, and other techniques are required for recognizing handwritten graphical symbols that are drawn in such these documents. In this paper we present a Dynamic Time Warping based method that outperforms the classical descriptors, being also invariant to scale, rotation, and elastic deformations typical found in handwriting musical notation.

1 Introduction

In the Graphics Recognition field, Optical Music Recognition (OMR) is a classical application area of interest, whose aim is the identification of music information from images of scores and their conversion into a machine readable format. In addition, the recognition of ancient manuscripts and their conversion to digital libraries can help in the diffusion and preservation of artistic and cultural heritage. In fact, most works have been done in the recognition of printed scores (see [1]), whereas few works can be found about the recognition of old handwritten ones (see [2], [3]). Working with old scores makes the recognition task more difficult due to paper degradation, the lack of a standard notation and the fact that staff lines are often handwritten. For those reasons, the preprocessing and segmentation phases must be adapted to this kind of scores.

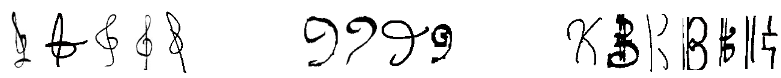
Our work is focused on the recognition of old handwritten scores (17th-19th centuries) so that these scores of unknown composers could be edited and published. The stages of the system are the following: first, the input gray-level scanned image is binarized with an adaptive binarization technique and morphological operations are used to filter the image and reduce noise. Afterwards, the image is deskewed using the Hough Transform method for detecting lines. Then, recognition and extraction of the staff lines (using median filters and a contour tracking process) and graphical primitives (using morphological operations) are performed. In this work we focus on the classification of the musical symbols using a Dynamic Time Warping based method.

2 Recognition of staff and primitives

The extraction of staff lines is difficult due to distortions in staff (lines present often gaps in between), and staff lines are rarely perfectly horizontal. This is caused by the degradation of old paper, the warping effect and the inherent distortion of handwritten strokes (staff lines are often drawn by hand). For those reasons, a more sophisticated process is required: After analyzing the histogram with horizontal projections of the image, detecting staff lines, a rough approximation of every staff line is performed using skeletons and median filters. Afterwards, a contour tracking algorithm is performed to follow every staff line and remove segments that do not belong to a musical symbol. Once we have the score without staff lines, vertical lines are recognized using median filters with a vertical structuring element, and filled headnotes are detected performing a morphological opening with elliptical structuring element. For further details, see [4].

3 Classification of handwritten musical symbols

Concerning the classification of old handwritten musical symbols (such as clefs, accidentals, time signature...), we state two main problems: the enormous variations in handwritten musical symbols and the lack of an standard notation in such these old scores. Thus, the classification process must cope with deformations and variations in writing style. Some tests have been performed using several descriptors for the clefs (Zernike moments, Zoning, ART, CSS), and several classifiers (KNN, Fisher, Adaboost, Support Vector Machine). In addition, unsupervised learning approaches have been tested using k-means inspired clustering methods to determine the number of significative classes. It has been proved that those descriptors do not reach good performance for old handwritten musical symbols, because there is no clear separability between classes (see Fig. 1).



(a) (b) (c)

Fig. 1. High variability of clefs appearance: (a) Treble clefs. (b) Bass clefs. (c) Alto clefs.

For those reasons, we are working in the research of other descriptors able to cope with the high variation in handwritting styles. The Dynamic Time Warping algorithm (DTW), described in [5], is commonly used in handwritten text recognition and speech recognition for computing the distance between two time series, optimizing the alignment because it minimizes a cumulative distance measure consisting of local distances between aligned samples. The main contribution of our work is to use the DTW approach for 2D shapes instead of 1D (in handwritten text it is used to align 1d sequences of pixels from the upper and lower contours).

Concretely, we are using this idea for the classification of old handwritten musical symbols, using some features of every symbol as as rough descriptors. First, for every column of the image we extract a set of features, consisting in the number of foreground pixels, the upper and lower profile, and the sum of pixels of every column region (3 files, 1 row). After normalizing these vectors, the DTW distance between *X(I)=*$x_1..x_M$ and *Y(J)=*$y_1..y_N$ is *D(M,N)*, calculated using:

$$D(i,j) = min \begin{cases} D(i,j-1), \\ D(i-1,j), \\ D(i-1,j-1), \end{cases} + d2(x_i, j_j) \tag{1}$$

$$d2(x_i, j_j) = \sum_{k=1}^{3} (f_k(I,i) - f_k(J,j))^2 \tag{2}$$

The length K of the warping path between X and Y (which can be obtained performing backtracking starting from *(M,N)*) biases the determined distance:

$$D(X,Y) = \sum_{k=1}^{K} d(x_{i_k}, y_{j_k}) \tag{3}$$

Finally, the matching cost is normalized by the length K of the warping path, otherwise longest symbols should have a bigger matching cost than the shorter ones:

$$MatchingCost(X,Y) = D(X,Y)/K \tag{4}$$

The DTW algorithm is efficient enough for comparing text or signals, but it will not work for comparing handwritten symbols because the slant of every symbol is usually different and can not be correctly computed. For those reasons, given two symbols to be compared, profiles for the DTW distance are extracted from different rotated versions, and the DTW distance will decide the best matching for both symbols. Once we have the minimum matching cost for the input symbol and the representatives of every class, the minimum distance will define the class where the input symbol belongs to.

4 Results and Conclusion

The DTW-based method for the classification of handwritten musical symbols has been evaluated using a database of clefs, which has been extracted from a collection of modern and old musical scores (19th century), containing a total of 2128 clefs between 24 different authors. In Fig. 2 the representatives chosen for each class are shown: one representative for treble and bass clefs and two representative alto clefs (because of its huge variability). Thus, only 4 comparisons are computed for classifying every input symbol. Table 4 shows the rates achieved using the classical Zernike moments and ART descriptors; and our DTW-based proposed descriptor, where a 95% rate is achieved.

(a) (b) (c) (d)

Fig. 2. Selected representative clefs: (a) Treble representative clef. (b) Bass representative clef. (c)(d) Two Alto representative clefs.

Method	Zernike moments	ART	DTW
Accuracy	65.07	72.74	95.81

Table 1. Classification of clefs: Recognition rates

Finally, one can see the outperform of our method in front of classical descriptors. In addition, it is invariant to scale, rotation and elastic deformations typical found in handwriting musical notation. Further work will be focused in the extension of the experiments into other handwritten musical symbols, such as accidentals and time signature.

Acknowledgements

This work has been partially supported by the project TIN2006-15694-C02-02.

References

1. D. Blostein, H. Baird, "A Critical Survey of Music Image Analysis", *Structured Document Image Analysis*, Eds. H. Baird, H. Bunke, and K. Yamamoto, Springer Verlag, 1992, pp 405-434.
2. J.C. Pinto, P. Vieira, J.M. Sosa, "A new graph-like classification method applied to ancient handwritten musical symbols", *International Journal of Document Analysis and Recognition (IJDAR)*, Vol. 6, Issue 1, 2003, pp 10-22.
3. N.P. Carter, "Segmentation and preliminary recognition of madrigals notated in white mensural notation", *Machine Vision and Applications*, Vol. 5, Number 3, 1995, pp 223-230.
4. A. Fornés, J. Lladós, and G. Sánchez, "Primitive segmentation in old handwritten music scores", in *Graphics Recognition: Ten Years Review and Future Perspectives*, ed. W. Liu and J. Llados, vol. 3926 of LNCS, Springer-Verlag, pp. 279-290, 2006.
5. T. Rath, R. Manmatha. "Word image matching using dynamic time warping". *In Proceedings of CVPR'03*, Vol. 2, 2003, pp 521-527.