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
{afornes,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 we work 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. It is a mature area of study, and lots of works have been done in the recognition of printed scores (see the survey written by Blostein in [1]).

Recently, the analysis of ancient documents has had an intensive activity, and the recognition of ancient musical scores is slowly being taken into account. In fact, the recognition of ancient manuscripts and their conversion to digital libraries can help in the diffusion and preservation of artistic and cultural heritage. It must be said that contrary to printed scores, few works can be found about the recognition of old handwritten ones (see [2], [3]). Working with old handwritten scores makes the recognition task more difficult: Firstly, and due to handwritten documents, one must cope with elastic deformations, the variability in the writer style, with variations in sizes, shapes and intensities, and increasing the number of touching and broken symbols. Secondly, working with old documents obviously increases the difficulties due to paper degradation, the frequent lack of a standard notation and the fact that staff lines are often handwritten. For those reasons, the preprocessing, segmentation and classification phases must be adapted to this kind of scores.

The specific processes required here belong to the field of Graphics Recognition, more than the field of Cursive Script Recognition(symbols are bidimensional). Symbol recognition is one of the central topics of Graphics Recognition. A lot of effort has been made in the last decade to develop good symbol and shape recognition methods inspired in either structural or statistic

W. Liu, J. Lladós, and J.-M. Ogier (Eds.): GREC 2007, LNCS 5046, pp. 51–60, 2008.

pattern recognition approaches. In [4], the state-of-the art of symbol recognition is reviewed. It must be said that the definition of expressive and compact shape description signatures is very important in symbol recognition, and has been an important area of study. In [5] the main techniques used in this field are reviewed. They are mainly classified in contour-based descriptors (such as polygonal approximations, chain code, shape signature, and curvature scale space) and region-based descriptors (such as Zernike moments, ART, and Legendre moments). A good shape descriptor should guarantee inter-class compacity and intra-class separability, even when describing noisy and distorted shapes. It has been proved that some descriptors which are robust with some affine transformations and occlusions in printed symbols, are not efficient enough for handwritten symbols. Thus, the research of other descriptors for elastic and non-uniform distortions are required, coping with variations in writing style.

In this paper we present our work in the recognition of old handwritten musical scores, which are from the 17th-19th centuries. The goal is not only the preservation of these old documents (see Fig.1 for an example), but also the edition and the diffusion of these unknown composers' compositions, which have not been published yet.

As it has been said above, handwritten recognition means dealing with elastic deformation. In Cursive Script Recognition it has been observed that the alignment (warping) of profiles of the shapes can cope with elastic deformations. For that reason, Dynamic Time Warping is a good solution to this problem. In fact, it has been successfully applied to handwritten text recognition in [12]. For

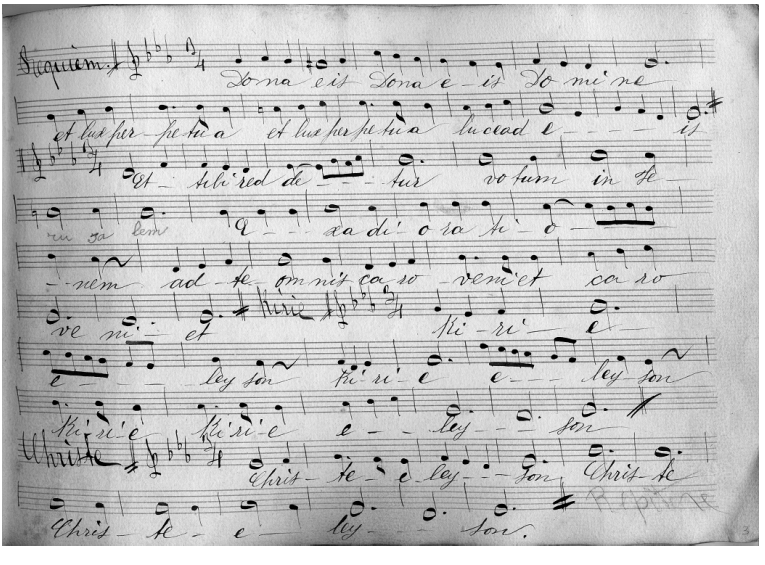


Fig. 1. Example of an old score (XIX century)

the classification of musical symbols we are applying the same concept, extending the method to two-dimensional graphical symbols. In addition, due to isolated symbols present in graphical documents, we must take into account the variations in rotation. For those reasons, the Dynamic Time Warping algorithm and the feature vectors have been adapted to 2D graphical symbols.

This paper is organized as follows: in section 2 the extraction of staff lines and the recognition of graphical primitives are presented. In section 3, the classification of handwritten musical symbols is fully described. It is performed using a Dynamic Time Warping based method, being invariant to rotation, scale and variations in writing style. In section 4 preliminary results over a database of musical symbols are shown. Finally, concluding remarks are presented.

2 Preprocessing, Staff Removal and Recognition of Graphical Primitives

For the sake of better understanding, we first briefly review our previous work for segmenting elements in the score (for further details, see [16]): First of all, 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.

The extraction of staff lines is difficult due to distortions in staff (lines present often gaps in between), and because of the fact that 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 (see Fig.2).

3 Classification of Handwritten Musical Symbols

Concerning the classification of old handwritten musical symbols, such as clefs, accidentals and 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 of the classical descriptors (such as Zernike moments, Zoning, ART) do not reach good performance for

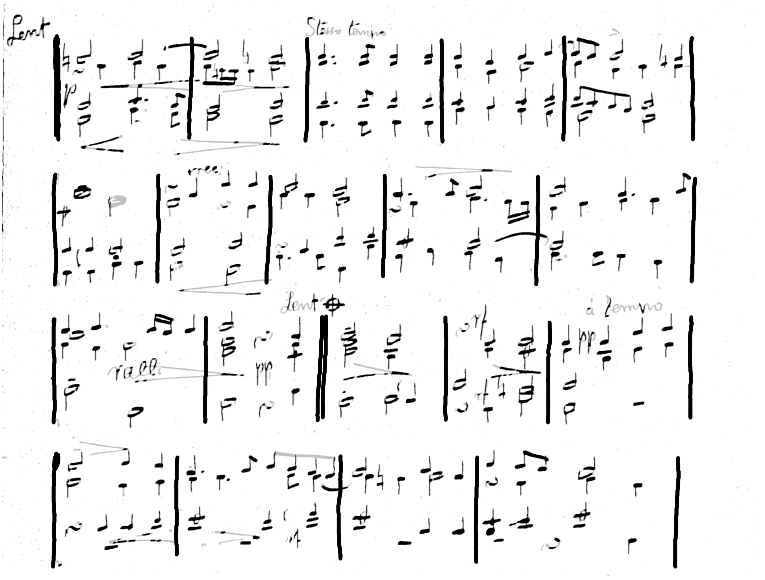


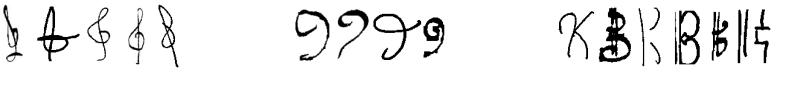
Fig. 2. Results from a section of 'Salve Regina' of the composer Aichinger: Filled headnotes and beams in black color. Bar lines are the thickest lines.

old handwritten musical symbols, because there is no clear separability between classes (the variability can be seen easily when we compare printed clefs with handwritten ones, see Fig. 3 and Fig. 4).

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

(a) (b) (c)

Fig. 3. Printed clefs: (a) Treble clef. (b) Bass clef. (c) Alto clef.



(a) (b) (c)

Fig. 4. High variability of handwritten clefs appearance: (a) Treble clefs. (b) Bass clefs. (c) Alto clefs.

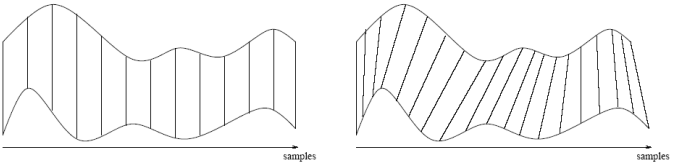


Fig. 5. Normal alignment and DTW alignment

(DTW) algorithm was first introduced by Kruskal and Liberman [6] for putting samples into correspondence in the context of speech recognition. DTW can warp the time axis, compressing it at some places and expanding it at others, avoiding the resample. Thus, it optimizes the best alignment (matching) between two samples, because it minimizes a cumulative distance measure consisting of local distances between aligned samples (see Figure 5).

Beside speech recognition, this technique has been widely used in many other applications, such as bioinformatics [7],[8], gesture recognition [9], data mining [10] and music audio recordings [11]. Rath and Manmatha have applied DTW to the handwritten recognition field [12], [13], coping also with the indexation of repositories of handwritten historical document. Also, Manmatha [14] has proposed an algorithm based on DTW for a word by word alignment of handwritten documents with their (ASCII) transcripts. Concerning online handwriting recognition, some work has also been done. Vuori [15] has also used a DTW based method for recognizing handwritten characters of several different writing styles. Concretely, the system retrieves a set of best matching allographic prototypes based on a query input character from an online handwriting system.

The main contribution of our work is to use a DTW approach for 2D shapes instead of 1D (in handwritten text it is used to align 1-dimensional 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 rough descriptors, and the Dynamic Time Warping algorithm (DTW) as the classifier technique used for clef matching: First, every image I is normalized, and for every column c (between 1 and w) of the image (where the width of the image is w pixels) we extract a set of features $X(I)=x_1..x_w$, where $x_c=(f_1,f_2,f_3..f_k)$ (see Fig.6), defined as:

- f_1 is the upper profile.
- f_2 is the lower profile.
- $f_3..f_k$ are the sum of pixels (zoning) of every column region (k-3+1 regions).

In handwritten text recognition (see [12]), features used are: f_1= sum of foreground pixels per column, f_2 = upper profile, f_3= lower profile, f_4= number of transitions from background to foreground. In our case of study, the sum of foreground pixels per column is not accurate enough (too many combinations of the same number of pixels per column can be found), so the column is divided

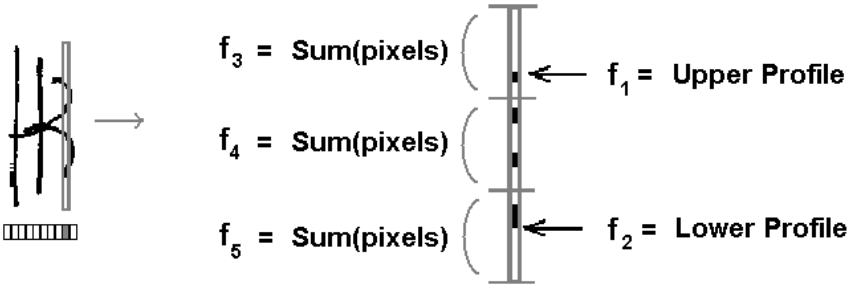


Fig. 6. Features extracted from every column of the image: f_1 = upper profile, f_2 = lower profile, $f_3..f_5$ = sum of pixels of the image of the three regions defined

in several regions and the sum of foreground pixels per region is computed (similar to a zoning). Concerning the number of transitions, when we work with old handwritten graphical symbols, the number of transitions and gaps can be very different from one symbol to another. For that reason, we have not included this feature.

After normalizing these vectors ($0 \leq f_s \leq 1$, s=1..k), the DTW distance between $X(I)=x_1..x_M$ and $Y(J)=y_1..y_N$ is $D(M,N)$, calculated using a dynamic programming approach:

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

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

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

$$D(X,Y) = \sum_{k=1}^{Z} d2(x_{i_k}, y_{j_k}) \tag{3}$$

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

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

The warping path is typically subject to several constraints, and once these conditions are satisfied, then the path that minimizes the warping cost is chosen:

– Boundary conditions: The warping path must start and finish in diagonally opposite corner cells of the matrix.

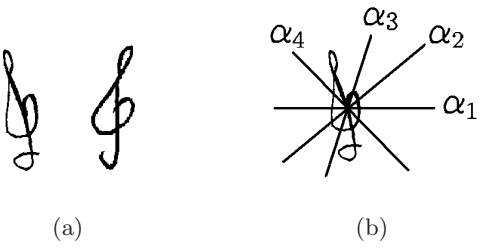


Fig. 7. Clefs: (a)Two treble clefs with different orientations. b)Some of the orientations used for extracting the features of every clef.

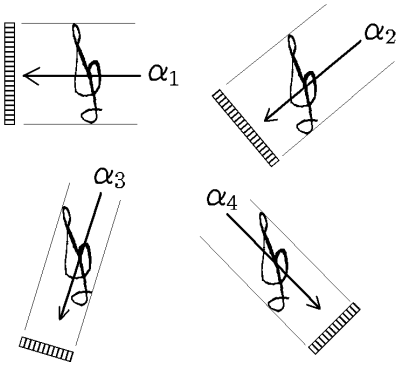


Fig. 8. Features extracted from some orientations ($\alpha_1..\alpha_4$)

- Continuity: The warping path must follow steps in adjacent cells (vertically, horizontally and diagonally adjacent cells).
- Monotonicity (monotonically increasing): This condition is for avoiding the matching from "going back in time".

The basic DTW algorithm will not work for comparing handwritten graphical symbols because the slant and the orientation of every symbol are usually different and it can not be easily computed (see Fig.7a). For that reason, given two symbols to be compared, profiles for the DTW distance are extracted from different orientations (see Fig.7b). Notice that the length of these profiles depends on the number of columns of the image, and varies from an orientation to another (see Fig.8).

Once we have the profiles for all the considered orientations, the DTW algorithm computes the matching cost between every orientation of the two symbols, and decides in which orientation these two symbols match in a better way.

In the classification stage, every input symbol is compared to the representatives of every class using this algorithm. The minimum distance will define the class where the input symbol belongs to.

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

Fig. 9. Selected representative clefs: (a) Treble representative clef. (b) Bass representative clef. (c)(d) Two Alto representative clefs.

4 Results

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. For every class, the representative chosen corresponds to the data sample with the minimum mean distance to the rest of samples from the same class. In Figure 9 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, where the 1-NN distance will decide which symbols belongs to each class.

The results from the DTW-based descriptor are compared with the classical Zernike moments and ART descriptors, because they are robust and invariant to scale and rotation. Table 1 shows the rates achieved using Zernike moments (number of moments = 7), ART (radial order = 2, angular order = 11) and our DTW-based proposed descriptor, where a 95% rate is achieved.

Table 1. Classification of clefs: Recognition rates of these 3 classes using 4 models

Method	Zernike moments	ART	DTW
Accuracy	65.07	72.74	95.81

An extension of these experiments has been performed including accidentals (sharps, naturals, flats and double sharps) in the musical symbol database. They are a total of 1970 accidentals drawn by 8 different authors. In Figure 10, one can see that some of them (such as sharps and naturals) can be easily misclassified due to their similarity. Contrary to double sharp, a double flat is just two flats drawn together, and for that reason double flats are not included in the accidentals' database.

Similarly to the experiments previously showed, we have chosen one representative for each class (see Fig. 11). The system will have now 8 models (4 clefs and 4 accidentals), and for every input symbol, 8 comparisons will be made.

Fig. 10. Accidentals (printed) in music notation

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

Fig. 11. Selected representative accidentals: (a) Sharp model. (b) Natural model. (c) Flat model. (d) Double sharp model.

Table 2. Classification of clefs and accidentals: Recognition rates of these 7 classes using 8 models

Method	Zernike moments	ART	DTW
Accuracy	43.97	52.26	89.55

Results are shown in Table 2, where the DTW-based proposed descriptor reaches a 89.55% classification rate, outperforming the results obtained by the other descriptors.

5 Conclusion and Future Work

In this paper we have presented an approach to classify musical symbols extracted from modern and old handwritten musical scores. This method is based in the Dynamic Time Warping method that has been extensively used in many applications. It is an extension of the approach used for handwriting text recognition, adapted to the 2D graphical symbols that are present in musical scores. One can see the outperform of our method in front of Zernike and ART descriptors. In addition, the method is invariant to scale, rotation and elastical deformations typically found in handwriting musical notation.

Further work will be focused on the extension of the experiments over several handwritten symbols databases, in order to show the robustness and scalability of the method. Finally, it must be said that the method could be improved using an expert system to learn every way of writing. Thus, the identification of the author in a musical score could be used for extracting knowledge information from a database, and helping in the classification stage.

Acknowledgements

This work has been partially supported by the spanish projects TIN2006-15694-C02-02 and CONSOLIDER-INGENIO 2010 (CSD2007-00018).

References

1. Blostein, D., Baird, H.: A Critical Survey of Music Image Analysis. In: Baird, H., Bunke, H., Yamamoto, K. (eds.) Structured Document Image Analysis, pp. 405–434. Springer, Heidelberg (1992)
2. Pinto, J.C., Vieira, P., Sosa, J.M.: A new graph-like classification method applied to ancient handwritten musical symbols. International Journal of Document Analysis and Recognition 6(1), 10–22 (2003)
3. Carter, N.P.: Segmentation and preliminary recognition of madrigals notated in white mensural notation. Machine Vision and Applications 5(3), 223–230 (1995)
4. Lladós, J., Valveny, E., Sánchez, G., Martí, E.: Symbol Recognition: Current Advances and Perspectives. In: Blostein, D., Kwon, Y.B. (eds.) GREC 2001. LNCS, vol. 2390, pp. 104–127. Springer, Heidelberg (2002)
5. Zhang, D., Lu, G.: Review of shape representation and description techniques. Pattern Recognition 37, 1–19 (2004)
6. Kruskal, J., Liberman, M.: The symmetric time-warping problem: from continuous to discrete. In: Sankoff, D., Kruskal, J. (eds.) Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison, pp. 125–161. Addison-Wesley Publishing Co., Reading (1983)
7. Aach, J., Church, G.: Aligning gene expression time series with time warping algorithms. Bioinformatics 17(6), 495–508 (2001)
8. Clote, P., Straubhaar, J.: Symmetric time warping, Boltzmann pair probabilities and functional genomics. Journal of Mathematical Biology 53(1), 135–161 (2006)
9. Gavrila, D.M., Davis, L.S.: Towards 3-D Model-based Tracking and Recognition of Human Movement. In: Bichsel, M. (ed.) Int. Workshop on Face and Gesture Recognition, pp. 272–277 (1995)
10. Keogh, E., Pazzani, M.: Scaling up dynamic time warping to massive datasets. In: Żytkow, J.M., Rauch, J. (eds.) PKDD 1999. LNCS (LNAI), vol. 1704, pp. 1–11. Springer, Heidelberg (1999)
11. Orio, N., Schwarz, D.: Alignment of monophonic and polyphonic music to a score. In: 2001 International Computer Music Conference, Havana, Cuba, pp. 155–158. International Computer Music Association, San Francisco (2001)
12. Rath, T., Manmatha, R.: Word image matching using dynamic time warping. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, WI, vol. 2, pp. 521–527 (2003)
13. Rath, T.M., Manmatha, R.: Lower-Bounding of Dynamic Time Warping Distances for Multivariate Time Series. Technical Report MM-40, Center for Intelligent Information Retrieval, University of Massachusetts Amherst (2003)
14. Kornfield, E.M., Manmatha, R., Allan, J.: Text Alignment with Handwritten Documents. In: First International Workshop on Document Image Analysis for Libraries, pp. 195–209. IEEE Computer Society, Washington (2004)
15. Vuori, V.: Adaptive Methods for On-Line Recognition of Isolated Handwritten Characters. PhD thesis, Helsinki University of Technology (2002)
16. Fornés, A., Lladós, J., Sánchez, G.: Primitive segmentation in old handwritten music scores, In: Liu, W., Lladós, J. (eds.) Graphics Recognition: Ten Years Review and Future Perspectives. LNCS, vol. 3926, pp. 279–290. Springer-Verlag (2006)