Writer Identification in Old Handwritten Music Scores

Alicia Fornés, Josep Lladós, Gemma Sánchez
Computer Vision Center - Dept. of Computer Science
Universitat Autònoma de Barcelona
Edifici O, 08193, Bellaterra, Spain
{afornes,josep,gemma}@cvc.uab.cat

Horst Bunke
Institute of Computer Science and Applied Mathematics
University of Bern, Neubrückstrasse 10
CH-3012 Bern, Switzerland
bunke@iam.unibe.ch

Abstract

The aim of writer identification is determining the writer of a piece of handwriting from a set of writers. In this paper we present a system for writer identification in old handwritten music scores. Even though an important amount of compositions contains handwritten text in the music scores, the aim of our work is to use only music notation to determine the author. The steps of the system proposed are the following. First of all, the music sheet is preprocessed and normalized for obtaining a single binarized music line, without the staff lines. Afterwards, 100 features are extracted for every music line, which are subsequently used in a k-NN classifier that compares every feature vector with prototypes stored in a database. By applying feature selection and extraction methods on the original feature set, the performance is increased. The proposed method has been tested on a database of old music scores from the 17th to 19th centuries, achieving a recognition rate of about 95%.

1. Introduction

Document analysis in historical documents has attracted growing interest in the last years, whose aim is the conversion of these documents into digital libraries, helping in the diffusion and preservation of artistic and cultural heritage. The presence of handwritten text, graphical illustrations or both in historical documents is common. Optical Music Recognition (OMR) is a classical area of interest of Document Image Analysis and Recognition (DIAR) that combines textual and graphical information. Contrary to printed scores (for a complete survey see [2]), few works have been done about the recognition of old handwritten ones (see [6], [16]). In addition to the preservation in digital format, the interest of aplying DIAR to historical handwritten musical scores is twofold. The first is the transcription of the score to a standard format, even machine readable like MIDI, while the second consists in the classification of the document in terms of the writer.

The identification of the author of a handwritten music score is still a challenge. In fact, many historical archives contain a huge number of sheets of musical compositions without information about the composer, and musicologists must work hard for identifying the writer (or the copier) of every sheet. For that reason, a system for writer identification in old music scores could help musicologists in such a task, which is time consuming and prone to errors.

Writer identification in handwritten text documents is a mature area of study (see [1], [12], [19], [20], [22]), whereas very few research has been done in identifying the writer of music scores. As far as we know, only one project (see [5], [9], [13]) has been performed about writer identification in music scores. The authors have developed a prototype that analyzes the music score and then extracts some features about structural information of the music symbols and notes. However this work is at a preliminary stage and no results have been published.

Most compositions in last centuries were sacred music, such as Requiem Masses, Stabat Mater, Glorias, Salve Regina or Matins. Consequently, they contain lyrics (text) for the chorus and the solists. In these scores, the writer identification methods for handwritten text documents could be applied for lyrics. However, the aim of our work is to evaluate the performance of writer identification methods extracting features only from music symbols. Moreover, our methodology will also be useful for writer identification in those music scores that contain no text, such as music scores for instruments.

In this paper we present a method for performing writer identification in musical scores, but avoid the recognition of the elements in the score. Some authors (see [7], [19], [20]) claim that writer identification in handwritten text documents can be performed without recognizing the words, i.e., with the meaning of the text being unknown. In the present paper, this assumption is extended to music scores. Consequently the system will be faster and more robust, avoiding the dependence on a good recognizer. In fact, we have adapted part of the writer identification approach described in [10] to old musical scores, where instead of letters of the alphabet, music notations are analysed.

The remainder of this paper is structured as follows. In the next section the preprocessing steps are presented, and in Section 3 feature extraction is described. In Section 4, an overview of the feature selection methods used in this paper is given. Experimental results are presented and discussed in Section 5. Finally, Section 6 concludes the paper and proposes future work.

2. Preprocessing

The preprocessing phase consists in binarizing the image, removing staff lines and normalizing the musical lines. Every output file contains the musical notation of one staff line. The process is described in the following subsections.

2.1. Binarization and Staff removal

The input gray-level scanned image (at a resolution of 300 dpi) is first binarized with the adaptive binarization technique proposed by Niblack [15]. Then, filtering and morphological operations are applied to reduce noise. Afterwards, the image is deskewed in order to make the recognition of staff lines easier. For this purpose, the Hough Transform method is used to detect lines and obtain the orientation of the music sheet. Then the image is rotated if necessary.

For writer identification, the staff lines are useful only if they are written by hand. In most of the music sheets of our database, however, they are printed. For that reason, staff lines are removed from the score. The extraction of staff lines (even if they are printed) is difficult because of paper degradation and the warping effect. For that reason, a robust system for detecting staffs is required, coping with distortions and gaps in staff lines. The steps for staff removal are the following. Firstly, a coarse staff approximation is obtained using horizontal runs as seeds to detect a segment of every staff line. This approximation is computed by applying median filters (with a horizontal mask) to the skeleton of the image. Remaining are only staff lines and horizontally-shaped symbols. Afterwards, staff lines are reconstructed, and each segment is discarded or joined with others according to its orientation, distance and area. Secondly, a contour tracking process is performed from left to right and right to left, following the best fitting path according to a given direction. In order to cope with gaps in staff lines and to avoid deviations (wrong paths) in the contour tracking process, the coarse staff approximation above described is consulted. Finally, those segments that belong to the staff lines (their width is similar to the average of the width of staff lines, which has been computed previously) are removed. For further details, see [8].

2.2. Normalization

The information about location of staff lines previously obtained is used for segmenting the music sheet into lines. Afterwards, the lines must be aligned with respect to a horizontal reference line. This step will be called normalization.

The normalization typically performed in handwritten text can not be applied here, because in musical scores, the height of every music line will vary depending on the melody of the composition. In music notation, notes are located upper or lower in the staff for reaching higher or lower frequency. Therefore, melodies with both treble and bass notes will result in a line with a larger height. This fact can be confusing for the writer identification system, which could wrongly identify heights of large extend in lines (melodies with bass and treble notes) as a typical feature of a specific writer. For that reason, the music notes must be rearranged with respect to a horizontal reference line. Thus, the normalization step computes the centroid of every connected component of the line, and uses this centroid for aligning the component with an horizontal reference line (see Fig.1).

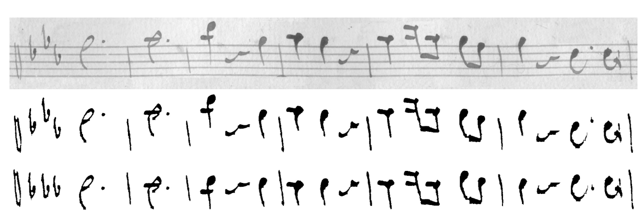

Figure 1. Preprocessing step: Original music line in gray scale, binarized music line (without staff lines), and normalized line.

3. Feature Extraction

Once the musical score is transformed into normalized handwritten individual music lines, 100 features are computed for every line. Previous work by Hertel and Bunke [10] was performed for writer identification in handwritten text documents. The idea is to use the same features, adapting them to music lines, within the specific normalization described in the previous section.

The 100 features proposed in [10] include basic features (such as slant and width of the writing), connected components, enclosed regions, lower and upper contour of the line and fractal features. More details are given below. For a full description we refer to [10] and [14].

3.1. Basic Features

The basic features taken into account are the following: the writing slant, the height of the main three zones and the width of the writing.

For obtaining the slant angle, the contour of the writing is computed and an angle histogram is created by accumulating the different angles along the contour. All angles are weighted by the length of the corresponding line. From the histogram, the mean and standard deviation are computed.

The three writing zones are called the UpperZone, the MiddleZone and the LowerZone. They are determined by the top line, the upper baseline, the lower baseline and the bottom line. To determine these lines, a horizontal projection of the music line is computed, and an ideal histogram with variable position of the upper baseline and the lower baseline is matched against this projection. Then, the following ratios (for avoiding absolute values) are used as features: U/M, U/L and M/L, where U is the height of the UpperZone, M is the height of the MiddleZone and L is the height of the LowerZone.

The width of the writing is obtained by selecting the row with most black-white and white-black transitions. Here, and for avoiding outliers, the median m_l of the lengths of every run is computed. Finally, this value is used for obtaining the ratio, M/m_l (where M is the height of the Middlezone), which will be used as a feature.

3.2. Connected Components

Some authors write musical notes in a continuous stroke while others break it up into a number of components. Thus, from every binary image of a line of music, connected components are extracted. Then, the average distance between two successive bounding boxes is computed. The system computes the average distance of two consecutive connected components and also the average distance between the elements belonging to the same connected component. Moreover, the average, median, standard deviation of the length of the connected components are used as features.

3.3. Enclosed Regions

Closed loops can be of circular, elliptical or rectangular shape, depending on the writing style. For that reason, features about the shape of the loops are useful and are added to the set of features. The loops are not analyzed directly. Instead, the blobs that are enclosed by a loop are computed by standard region growing algorithm. The first feature is the average of the form factor f, taken over all blobs of one line. If A is the area of the blob under consideration and l is the length of its boundary, the form factor f is computed as:

$$f = \frac{4A\pi}{l^2} \tag{1}$$

The second feature measures the roundness r of an object as follows:

$$r = \frac{l^2}{A} \tag{2}$$

Finally, the average over all blobs and the average size of the blobs in a line are taken as features.

3.4. Lower and Upper Contour

A visual analysis of the upper and lower contours of the music lines reveals that they differ from one writer to another. Some writings show a rather smooth contour whereas others are pointed with more peaks, being useful information for writer identification.

For selecting the lower and the upper contour of a line, gaps must be removed, and discontinuities in the y-axis are eliminated by shifting these elements along the y-axis. Once the continuous lower and upper contour (called characteristic contours) are obtained, the following features are extracted: slant of the characteristic contour (obtained through linear regression analysis), the mean squared error between the regression line and the original curve, the frequency of the local maxima and minima on the characteristic contour (if m is the number of local maxima and l is the number of local minima, then the frequency of local maxima is m/l and the frequency of local minima is l/m), the local slope of the characteristic contour to the left of a local maximum within a given distance, and the average value taken over the whole characteristic contour. The same features are computed for the local slope to the right of a local maximum, and the same for local minima to the right and to the left.

3.5. Fractal Features

The idea proposed in [3],[4] is to measure how the area A of a handwritten line grows when a morphological dilation operation is applied on the binary image. The line is first thinned, and the dilation is performed using different kernels (disks of radius η for information invariant to rotation).

For each of this kernels, the area $A(X_\eta)$ of the dilated writing X_η is measured. The fractal dimension $D(X)$ is defined by:

$$D(X) = \lim_{\eta \to 0}(2 - \frac{\ln A(X_\eta)}{\ln \eta}) \tag{3}$$

Then, we obtain the evolution graph plotting the behaviour of y over x (see Fig.2):

$$x = \ln \eta \tag{4}$$

$$y = \ln A(X_\eta) - \ln \eta \tag{5}$$

Afterwards, this function is approximated by three straight lines (see Fig.2). The points $p_1, ..., p_4$ are found by minimizing the square error between the three line segments and the points of the evolution graph. Finally, the slopes of these three characteristic straight line segments are computed and used as features.

In addition to three disks kernels, 18 ellipsoidal kernels are used for getting information about the rotation in the writting style. These ellipses are defined with increasing the length of the ellipse's two main axes and the rotation angle. Thus, a total of 63 (=21x3) features are extracted.

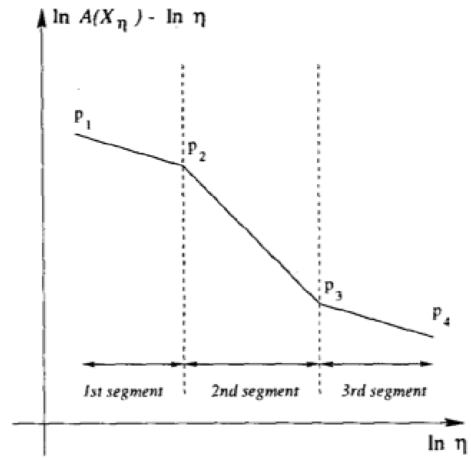

Figure 2. Fractals: Approximation of the evolution graph by three straight lines (from [14]).

4. Feature Selection

In [21] the suitability of the 100 features described in Section 3 has been analyzed, because some of them could be unnecessary or even redundant. The goal of feature selection is to find the best subset of features that perform better than the original ones, and also, results in a more efficient classifier.

There are two main groups of methods for feature selection: feature set searching and linearly combining features for getting lower dimensionality. These two groups are described in next subsections.

4.1. Feature Set Search

The main idea of Feature Set Searching methods is to find the best subset of features for classification. In [11] and [17] four techniques are described: Sequential Forward Search (SFS), Sequential Backward Search (SBS), Sequential Floating Forward Search (SFFS), Sequential Floating Backward Search (SFBS).

SFS starts with an empty set of features, and at each step one single feature is added to the set. The feature chosen is the best classifying feature from the remaining set of features. Contrary, SBS starts with the full set of features, and removes one feature so that the new reduced set of features yields a higher writer identification rate. SFBS and SFFS are an improvement of SFS and SBS, adding the ability to do backtracking. The set of features can be incremented or reduced by one feature at each time, changing dynamically the number of features in the set, thus floating up and down. SFBS starts with the empty set of features, whereas SBS starts with the full set of features.

4.2. Feature Combination

Principal Component Analysis (PCA) and Multiple Discriminant Analysis (MDA) (see [18]) are two of the main classical methods for reducing dimensionality. They linearly combine the original features and then project the new ones onto a space of lower dimensionality.

PCA seeks a projection that best represents the data. It first computes the mean and variance of all feature vectors and normalizes them. Then the covariance matrix and its eigenvectors and eigenvalues are computed. The eigenvectors corresponding to the larges eigenvalues are retained and the input vectors are projected onto the subspace defined by these eigenvectors. The vectors of this lower dimensional space are then used in the K-Nearest Neighbour classifier.

MDA is the second method used, which seeks a projection that best separates the data. It is an extension of Fisher's linear discriminant analysis from a two-class to a c-class classification problem, projecting high-dimensional data onto a line and performing classification in this one-dimensional space. The projection maximizes the distance between the means of the two classes and simultaneously minimizes the variance within each class. The resulting vectors are also used in the K-NN classifier.

5. Experimental Results

We have tested our method with sets of 25 music lines each from one out of seven different writers, obtaining a database with a total of 175 music lines. These music lines are extracted from a collection of music scores of the 17th, 18th and 19th centuries, which have been obtained from two archives in Catalonia (Spain): the archive of Seminar of Barcelona and the archive of Canet de Mar. An example of an old score can be seen in Figure 3.

Figure 3. Example of an old score of the composer Casanoves.

The music lines are obtained through the preprocessing steps described above, and the vector of 100 features is computed for every music line. The classification has been performed using a 5-Nearest Neighbour (5-NN) classifier based on Euclidean distance. The chosen value k=5 has been empirically determined to be the optimal choice. In the experiments, we have used 5 test subsets, randomly chosen, containing 3 music lines per writer. Thus, every test set of 21 files is classified using a training data set of 154 files.

For the SFS, SBS, SFFS and SFBS experiments, wrappers are used as objective function, where one of the five subsets is used as the test set and the others as prototypes in the 5-NN classifier. To evaluate the fitness of a selected feature subset, iteratively three subsets are used in the classifier and the remaining set is used to measure the fitness of the feature subset under consideration. Once the algorithm finds the best feature subset, the fifth subset is used for the final writer identification rate.

In the PCA and FDA experiments, the optimal dimension of the transformed feature subspace is experimentally estimated as follows. Iteratively, three of the four subsets of the training set are used as prototypes in the 5-NN classifier, and the writer identification rate is calculated for a given dimension on the fourth set. The average of the four rates is computed, and the dimension which produces the highest rate is selected. Finally, we use this dimension for computing the classification rate on the test set, using the four training subsets.

Table 1. Classification Results: Writer identification rates using all 100 features and also some subsets of features, selected by groups of features.

Set of Features	W.I.Rate
All 100 features	79%
All - Basic	81%
All - Contours	77%
All - Component	81%
All - Fractal	73%
All - Regions	81%
Only Fractal	71%
All - {Basic, Contours} + Slant	73%
All - {Basic, Contours, Regions} + Slant	75%
All - {Basic, Contours, Regions} + {Slant, Area}	73%

In Table 1 the classification results for various sets of features are shown. The first row shows the writer identification rate using all 100 features, which is about 79%. The next five rows show the rates when we remove a group of features from the original set of 100 features: all features except the basic ones, all features except contours, all features except connected component features, all features except fractal features and all features except enclosed regions features. The next row shows the writer identification rates using only the 63 fractal features. The last three rows show the classification rates when selecting some random groups of features. In the eighth row the set is composed by all features except the basic and contours ones, but adding the slant (which belongs to the basic features). In the ninth row the set is composed by the same features than the eighth one, but removing the enclosed regions features. In the last row, the set is composed by the same features than in the ninth row, but adding the area (which belongs to the enclosed regions set of features). These sets of features have been randomly chosen, showing that one could look for a subset of the initial 100 features for improving the final writer identification rate. Obviously, the selection of features in a intuitive and manual way does not reach a significant improvement (from the baseline rate of 79% to 81%). For that reason, the feature selection methods are required.

In Table 2 results of feature selection algorithms are shown. The first row again shows the baseline rate, where all 100 features are used for the classification. The next ones show the writer identification rates using PCA (79%) and MDA (63%) methods, which do not improve the final classification rate. In fact, MDA reaches a rate below than the original 100 features. In MDA the maximum number of features is equal to the number of classes minus one. So, this result is quite comprehensible because the maximum number of features is limited to six in this application, being too small for achieving good rates. The last four rows show the results using SFS, SBS, SFFS and SFBS feature set search methods. They reach significant improvements over the baseline (all are over 93%). Among them, SFBS reaches an identification rate of 95% using only 45 features, being the best writer identification rate of all. This fact shows that there are many dependent or irrelevant features in the original feature set, giving us the possibility to select a subset for improving the results in this database. Notice that these selected features are specific to this database, and the results could potentially be quite different for other datasets.

Table 2. Classification Results: Writer identification rates using Feature Set Search and Feature Combination methods.

Experiment	N. of Features	W.I.Rate
All Features	100	79%
PCA	88	79%
MDA	6	63%
SFS	16	93%
SBS	16	94%
SFFS	41	93%
SFBS	45	95%

6. Conclusions and Future Work

In this paper we have presented a method for writer identification in musical scores. The steps of the system are the following. In the preprocessing step, the image is binarized, de-skewed, staffs and removed and the lines of music are normalized. Afterwards, 100 features (slant, connected components, enclosed regions, upper and lower contours, and fractals) are computed. Finally, the classification is performed using the k-Nearest Neighbour method. One can see how classification rates for this specific database are significantly increased using feature selection methods.

The work is still at an early stage, but we have obtained very high classification rates. Further work will be focused on increasing the database, using more classifiers (such as Support Vector Machines or Neural Networks), experimenting with other feature selection methods, and adding specific features for musical notation to the current set of features.

Acknowledgements

The authors would like to thank Prof. Josep Maria Gregori Cifré from Art Department of UAB for his help in accessing to old resources of archive of Seminar of Barcelona and the archive of Canet de Mar. Furthermore, we would also like to thank Vivian Kilchherr and Dr. Andreas Schlapbach for providing guidance and support in the feature extraction and search experiments.

This work has been partially supported by the Spanish projects TIN2006-15694-C02-02 and CONSOLIDER-INGENIO 2010 (CSD2007-00018) and catalan scholarship 2007 BE-1 00086.

References

[1] A. Bensefia, T. Paquet, and L. Heutte. A writer identification and verification system. *Pattern Recognition Letters*, 26(13):2080–2092, 2005.

[2] D. Blostein and H. S. Baird. A critical survey of music image analysis. In H. Baird, H. Bunke, and K. Yamamoto, editors, *Structured Document Image Analysis*, pages 405–434. Springer-Verlag, Heidelberg, 1992.

[3] V. Bouletreau, N. Vincent, R. Sabourin, and H. Emptoz. Synthetic parameters for handwriting classification. In *Proceedings of the 4th International Conference on Document Analysis and Recognition (ICDAR)*, pages 102–106, Washington, DC, USA, 1997. IEEE Computer Society.

[4] V. Bouletreau, N. Vincent, R. Sabourin, and H. Emptoz. Handwriting and signature: One or two personality identifiers? In *Proceedings of 14th International Conference on Pattern Recognition (ICPR)*, volume 2, pages 1758–1760, 1998.

[5] I. Bruder, T. Ignatova, and L. Milewski. Integrating knowledge components for writer identification in a digital archive of historical music scores. In *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital libraries (JCDL)*, pages 397–397, New York, NY, USA, 2004. ACM.

[6] N. P. Carter and R. A. Bacon. Automatic recognition of printed music. In H. Baird, H. Bunke, and K. Yamamoto, editors, *Structured Document Image Analysis*, pages 169–203. Sringer-Verlag, 1992.

[7] S.-H. Cha and S. N. Srihari. Writer identification: Statistical analysis and dichotomizer. In *Proceedings of the Joint IAPR International Workshops on Advances in Pattern Recognition*, pages 123–132, London, UK, 2000. Springer-Verlag.

[8] A. Fornés, J. Lladós, and G. Sánchez. Primitive segmentation in old handwritten music scores. In W.Liu and J.Lladós, editors, *Graphics Recognition: Ten Years Review and Future Perspectives*, volume 3926 of LNCS, pages 279–290. Springer, 2006.

[9] R. Göcke. Building a system for writer identification on handwritten music scores. In *Proceedings of the IASTED International Conference on Signal Processing, Pattern Recognition, and Applications (SPPRA)*, pages 250–255, Rhodes, Greece, 30 June – 2 July 2003.

[10] C. Hertel and H. Bunke. A set of novel features for writer identification. In *Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 679–687, 2003.

[11] J. Kittler. Feature set search algorithms. In C.H.Chen, editor, *Pattern Recognition and Signal Processing*, 1978.

[12] F. Leclerc and R. Plamondon. Automatic signature verification: The state of the art. *International Journal of Pattern Recognition and Artificial Intelligence (PRAI)*, 8(3):643–660, 1994.

[13] N. Luth. Automatic identification of music notations. In *Proceedings of the Second International Conference on WEB Delivering of Music (WEDELMUSIC)*, 2002.

[14] U. Marti, R. Messerli, and H. Bunke. Writer identification using text line based features. In *Proceedings of the 6th International Conference on Document Analysis and Recognition (ICDAR)*, pages 101–105, 2001.

[15] W. Niblack. *An Introduction to Digital Image Processing*. Prentice Hall, 1986.

[16] J. Pinto, P. Vieira, and J. Sosa. A new graph-like classification method applied to ancient handwritten musical symbols. *International Journal of Document Analysis and Recognition (IJDAR)*, 6(1):10–22, 2003.

[17] P.Pudil, J.Novovicová, and J.Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15:1119–1125, 1994.

[18] R.O.Duda, P.E.Hart, and D.G.Stork. *Pattern Classification*. Wiley Interscience, 2001.

[19] H. Said, T. Tan, and K. Baker. Personal identification based on handwriting. *Pattern Recognition*, 33(1):149–160, January 2000.

[20] A. Schlapbach and H. Bunke. Off-line writer identification and verification using gaussian mixture models. In S. Marinai and H. Fujisawa, editors, *Machine Learning in Document Analysis and Recognition*, volume 90 of *Studies in Computational Intelligence*, pages 409–428. Springer, 2008.

[21] A. Schlapbach, V. Kilchherr, and H. Bunke. Improving writer identification by means of feature selection and extraction. In *Proceedings of the 8th International Conference on Document Analysis and Recognition (ICDAR)*, pages I: 131–135, 2005.

[22] L. Schomaker and M. Bulacu. Automatic writer identification using connected-component contours and edge-based features of uppercase western script. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):787–798, June 2004.