Agència
de Gestió d'Ajuts
Universitaris
i de Recerca

Memòria justificativa de recerca de les convocatòries BCC, BE, BP, CTP-AIRE, DEBEQ, FI, INEFC, NANOS i PIV

La memòria justificativa consta de les dues parts que venen a continuació:
1.- Dades bàsiques i resums
2.- Memòria del treball (informe científic)

Tots els camps són obligatoris

1.- Dades bàsiques i resums

Nom de la convocatòria

BE

Llegenda per a les convocatòries:

BCC	Convocatòria de beques per a joves membres de comunitats catalanes a l'exterior
BE	Beques per a estades per a la recerca fora de Catalunya
BP	Convocatòria d'ajuts postdoctorals dins del programa Beatriu de Pinós
CTP-AIRE	Ajuts per accions de cooperació en el marc de la comunitat de treball dels Pirineus. Ajuts de mobilitat de personal investigador.
DEBEQ (Modalitat A3)	Beques de Cooperació Internacional i Desenvolupament
FI	Beques predoctorals per a la formació de personal investigador
INEFC	Beques predoctorals i de col·laboració, dins de l'àmbit de l'educació física i l'esport i les ciències aplicades a l'esport
NANOS	Beques de recerca per a la formació en el camp de les nanotecnologies
PIV	Beques de recerca per a professors i investigadors visitants a Catalunya

Títol del projecte: ha de sintetitzar la temàtica científica del vostre document.
Anàlisi i Identificació de l'escriptor de partitures musicals antigues

Dades de l'investigador o beneficiari

Nom	Cognoms
Alicia	Fornés Bisquerra

Correu electrònic
afornesb@yahoo.es

Dades del centre d'origen

Universitat Autònoma de Barcelona

Número d'expedient
2007 BE-1 0008

Paraules clau: cal que esmenteu cinc conceptes que defineixin el contingut de la vostra memòria.
Anàlisi de Documents
Reconeixement de Gràfics
Identificació de l'escriptor
Reconeixement Òptic Musical
Documents Antics

Data de presentació de la justificació
17-02-09

Nom i cognoms i signatura
del/de la investigador/a

Visttiplau del/de la responsable de la
sol·licitud

Resum del projecte: cal adjuntar dos resums del document, l'un en anglès i l'altre en la llengua del document, on s'esmenti la durada de l'acció

Resum en la llengua del projecte (màxim 300 paraules)

El treball realitzat durant els tres mesos d'estada a la Universitat de Berna (Suïssa) consisteix en el desenvolupament d'un sistema que permeti la identificació de l'escriptor de partitures musicals antigues.
La identificació de l'escriptor consisteix en determinar l'autor d'un document manuscrit d'entre un conjunt d'escriptors. L'objectiu és emprar només notació musical per determinar l'autor (encara que hi ha partitures que contenen també text). S'han desenvolupat dos mètodes per a la identificació: el primer proposa extreure característiques de cada linia musical (és a dir, de cada pentagrama), i el segon extreu característiques texturals de textures de símbols musicals.
Primerament, la partitura és preprocessada, binaritzant-la i eliminant les línies de pentagrama.

El primer mètode extreu 100 característiques de cada línia musical: mesures bàsiques (com és la inclinació i el gruix de l'escriptura), components connexes, regions tancades, contorns del perfil superior i inferior, i característiques fractals.

El segon mètode genera 4 diferents tipus de textura a partir dels símbols musicals. Cada mètode empra diferents variacions espaials per combinar els símbols musicals i generar una textura. Després, filtres de Gabor i Matrius de co-ocurrència de nivells de grisos (GSCM) son emprats per extreure característiques.

En ambdós casos la classificació es realitza emprant un classificador k-NN basat en la distància euclidea.
Els mètodes han estat provats sobre una base de dades de partitures antigues (segles XVII-XIX), aconseguint encoratjadors resultats d'identificació.

Resum en anglès (màxim 300 paraules)

The research performed during three months in the University of Bern (Switzerland) consists in the development of a method for writer identification in old handwritten music scores.

Writer identification consists in determining the writer of a piece of handwriting from a set of writers.
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.
We have developed two approaches for writer identification in old handwritten music scores. The methods proposed extract features from every music line, and also features from a texture image of music symbols. First of all, the music sheet is first preprocessed for obtaining a binarized music score without the staff lines.

In the first approach, 100 features are extracted from every music line: basic measurements (slant, width of the writing...), connected components, enclosed regions, contours of the upper and lower profile and fractal features.

In the second approach, four different methods for generating texture images from music symbols are applied. Every approach uses a different spatial variation when combining the music symbols to generate the textures. Afterwards, Gabor filters and Grey-scale Co-ocurrence matrices are used to obtain the features.

Resum en anglès (màxim 300 paraules) – continuació -.

The classification is performed using a k-NN classifier based on Euclidean distance.
The proposed method has been tested on a database of old music scores from the 17th to 19th centuries, achieving encouraging identification rates.

2.- Memòria del treball (informe científic sense limitació de paraules). Pot incloure altres fitxers de qualsevol mena, no més grans de 10 MB cadascun d'ells.

Veure pdf adjunt.

On the Writer Identification in Old Handwritten Music Scores

Alicia Fornés

Contents

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. Optical Music Recognition (OMR) is a classical area of interest of Document Image Analysis and Recognition (DIAR) that combines textual and graphical information. 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. 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.

Writer identification in handwritten text documents is a mature area of study (see [1], [10], [16], [17], [19]), whereas very few research has been done in identifying the writer of music scores. As far as we know, only one project (see [4], [7], [11]) 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, containing lyrics (text) for singers. 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 two methods for performing writer identification in musical scores, which avoid the recognition of the elements in the score. Some authors (see [5], [16], [17]) 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 [8] and [16] to old musical scores, where instead of letters of the alphabet, music notations are analysed.

The remainder of this document is structured as follows. In the next section the preprocessing steps are presented, and in Section 3 and 4 the two methods for feature extraction are described. In Section 5, experimental results are presented. Finally, Section 6 concludes the paper and proposes future work.

2 Preprocessing

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 [13]. 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 and removing every segment of the staffs is developed, which uses median filters and contour tracking. For further details, see [6].

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).

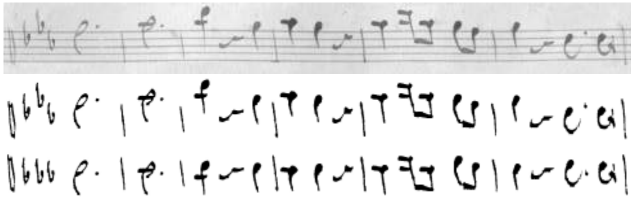

Figure 1: Preprocessing step: Original music line in gray scale, binarized music line (without staff lines), and normalized line.

3 Extraction of line features

The first method proposed extracts 100 features for every music text line. For a full description we refer to [8] and [12].

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 and the roundness r are computed as:

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

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 [2],[3] 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:

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

Then, we obtain the evolution graph plotting the behaviour of y over x:

$$x = \ln \eta, \qquad y = \ln A(X_\eta) - \ln \eta \tag{3}$$

Afterwards, this function is approximated by three straight lines. 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.

4 Extraction of texture features

The second method proposed extracts texture features from music textures.

4.1 Creation of textures

Taking the music lines in the preprocessing step, one must create textures for extracting features. There are four different approaches for obtaining these textures. In Figure 2 the four different textures from two writers are shown. The first kind of texture (Fig. 2(a),(b)) is called Basic texture, which consists in taking altogether the text lines obtained in the preprocessing step, without any processment. The second one (Fig. 2(c),(d)) is called TextLine texture, which consists in taking randomly music symbols and putting them in a reference line, with the same inter-symbol distance. In this way, if the music score contains a group of the same kind of music symbol (i.e. quarters or rests), they will be randomly distributed over the texture, obtaining a more heterogeneous one. The third one (Fig. 2(e),(f)) is called Aspect Ratio texture, which consists in taking the idea of TextLine texture, but making all the symbols of equal size. For every symbol that must be resized, its aspect ratio will be maintained. The last one is called Resize texture, which consist in the same idea of Aspect Ratio texture, but without the preservation of the aspect ratio in the resize process.

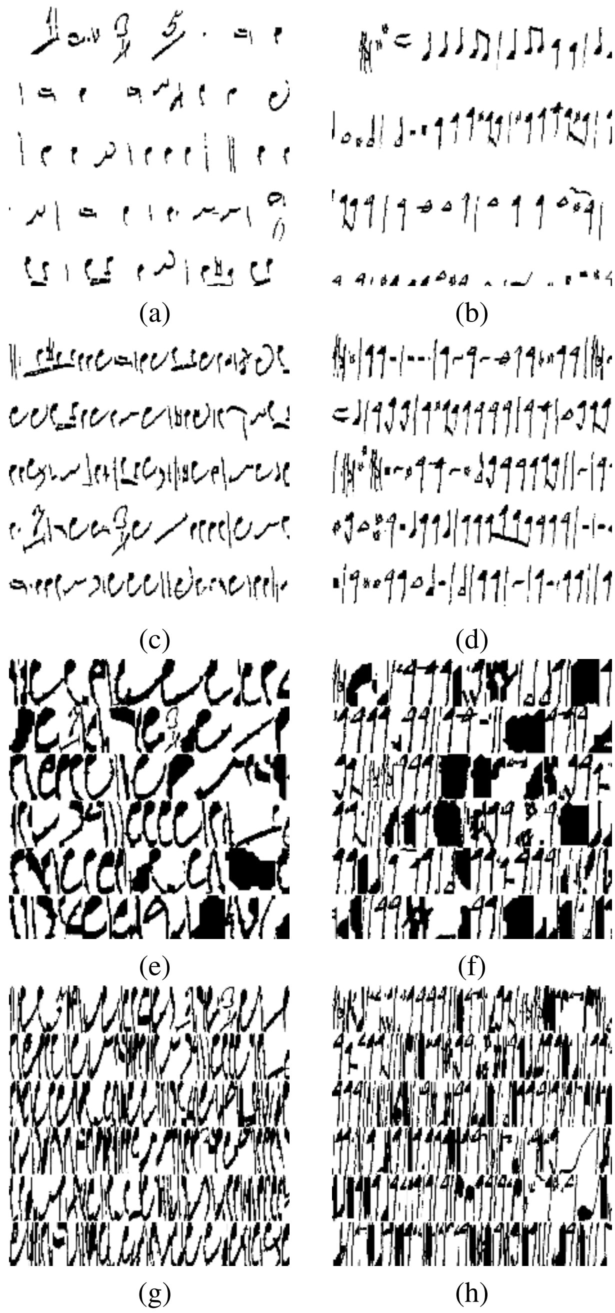

Figure 2: Textures created from music symbols of two authors. The first column shows textures from Clausell, the second column corresponds to textures from Milans: (a)-(b) Basic, textures (c)-(d) TextLine textures, (e)-(f) Aspect ratio textures, (g)-(h) Resize textures.

4.2 Feature extraction from textures

Once we have the music textures, textural features are extracted. Some work by Said [16] has been performed using texture features for writer identification in handwritten text. The idea is to use the same features, adapting them to music textures.

4.2.1 Gabor features

The multi-channel Gabor filtering technique is based in the psycophysical findings that affirm that the processing of pictorial information in the human visual cortex involves a set of parallel and quasi-independent cortical channels, which can be modelled by bandpass filters. The two Gabor filters are of opposite symmetry and are computed as:

$$\begin{cases} h_e(x,y;f,\theta) = g(x,y)cos(2\pi f(xcos\theta + ysin\theta)) \\ h_o(x,y;f,\theta) = g(x,y)sin(2\pi f(xcos\theta + ysin\theta)) \end{cases} \tag{4}$$

Afterwards, the Fourier transform (FFT) of the filters are computed as:

$$\begin{cases} q_o(x,y) = FFT^{-1}\left[P(u,v)H_o(u,v)\right] \\ q_e(x,y) = FFT^{-1}\left[P(u,v)H_e(u,v)\right] \end{cases} \tag{5}$$

where $P(u,v)$ is the Fourier Transform of the input image $p(x,y)$ and $H_o(u,v)$ and $H_e(u,v)$ are the Fourier Transform of the filters $h_o(x,y;f,\theta)$ and $h_e(x,y;f,\theta)$ respectively. Finally, we obtain the combination of the two filters, and a single value at each pixel is obtained:

$$q(x,y) = \sqrt{q_e^2(x,y) + q_o^2(x,y)} \qquad (6)$$

For the computation of features, the two parameters used are the radial frequency with values $f = \{4, 8, 16, 32\}$; and the orientation with values $\theta = \{0^o, 45^o, 90^o, 135^o\}$. The output corresponds to 4x4 = 16 images. Extracting the mean and the standard deviation we obtain a total of 16x2 = 32 Gabor features.

4.2.2 GSCM features

The Grey-Scale Co-ocurrence Matrices (GSCM) are typically used for describing grey scale textures. For an image with N grey levels, each GSCM is a matrix NxN, with a distance d and an angle θ, where $GSCM_{d,\theta}(a,b)$ corresponds to the number of pairs $(Pixel1, Pixel2)$ which $Pixel1$ is of grey color a, $Pixel2$ is of grey color b, and Pixel1 and Pixel2 are separated a distance d and angle θ. Whereas GSCM are of a high computational cost for grey images, they are fast to compute with binary images, because there are only two grey values: black and white.

The parameters used in our method are the distance d with values $d = \{1, 2, 3, 4, 5\}$; and the orientation $\theta = \{0^o, 45^o, 90^o, 135^o\}$. The output corresponds to 20 matrices of dimension 2x2, with 3 independent values (there is diagonal symmetry). In total we obtain 20x3 = 60 features.

5 Experimental Results

5.1 Results using line features

We have tested the first 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.

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 k-Nearest Neighbour classifier based on Euclidean distance. 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. In Table 1 the classification results for different values of k-NN are shown, where 3-NN reaches better writer identification rates (83.8%).

Table 1: Writer identification rates using all 100 features and different values of k-NN.

k-NN	W.I.Rate
1-NN	80.0%
3-NN	83.8%
5-NN	79.0%
7-NN	79.0%

5.2 Results using texture features

We have tested the second method with 66 music pages from 6 different writers, where every writer has written 11 pages. Three textures are created per music page, obtaining a database of 6x33 = 198 music textures. These pages are extracted from the same collection of music scores explained in the previous subsection.

As in the previous experiments, we have also used 5 test subsets, randomly chosen, containing one page per writer. This means that all the three music textures obtained from every page are used in the test set. Thus, every test set of 6x3= 18 files is classified using a training data set of 198-18 = 180 files.

In Table 2 the classification results for different values of k-NN are shown. Resize textures obtain better writer identification rates (83%) using Gabor features, whereas GSCM obtain the best recognition rates (over 90%) in both TextLine and Resize textures. Notice that the textures extracted using the Aspect Ratio technique obtain in both cases the worst writer identification rates. This fact can be explained because of the big black spots that appear in the textures, confusing the system.

Table 2: Writer identification rates using Gabor and GSCM features for different values of k-NN.

Experiment	Basic	TextLine	A.Ratio	Resize
Gabor 1-NN	65%	69%	61%	83%
Gabor 3-NN	63%	68%	55%	78%
GSCM 1-NN	81%	93%	63%	92%
GSCM 1-NN	82%	93%	65%	92%

6 Conclusions and Future Work

We have presented two methods for writer identification in musical scores. The first one uses line features whereas the second one uses textural features. The work is still at an early stage, but we have obtained high classification rates. Further work will be focused on increasing the database and adding specific features for musical notation to the current set of features.

References

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

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

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

[10] 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.

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

[12] 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.

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

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

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

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

[17] 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.

[18] 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.

[19] 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.