BUILDING A SYSTEM FOR WRITER IDENTIFICATION ON HANDWRITTEN MUSIC SCORES

Roland Göcke*
Dept. Human-Centered Interaction & Technologies
Fraunhofer Institute of Computer Graphics, Division Rostock
Rostock, Germany
email: Roland.Goecke@rostock.igd.fhg.de

ABSTRACT

17th and 18th century music scores were copied and distributed in a manual way. Music historians are interested in how the compositions were distributed or in other words, who copied the compositions when and where. Such information may also help to determine the composer when a piece of unknown origin is found. In this paper, we present ongoing work on the development of a software system to analyse such documents automatically and to aid the musicologists in their task to register handwritten music scores. In particular, we focus on the application and adaptation of image processing methods to separate music symbols for the identification task from irrelevant elements.

KEY WORDS

Image Processing, Music Scores, Handwriting Identification

1 Introduction

In past centuries, before the introduction of printed music scores, compositions were copied by hand. The copies were then passed around and copied further which was the way of spreading a composer's work. Such music pieces were often collected by the nobelity and their collections form a valuable source of information and cultural heritage. While the composer of a piece of music is often known, information about who copied the scores where and when is often unavailable. However, this information is important for the work of music historians as they try to re-establish the information lost in time and to register the music scores found in archives. In addition, information about the time and place of creation of a copy of a music piece can also deliver important information on the composer in the case of a piece of unknown origin.

The task of a musicologist is similar to that of a criminologist in terms of using handwritings and other available information. However, in this case the handwritten symbols are music score notations instead of letters of the alphabet. A strict definition of music symbols certainly did not exist in the 17th and 18th century, so that differences in the notation of the same symbol exist between different composers and copyists. Furthermore, just like handwritten characters, handwritten music symbols are produced in a personal manner. Musicologists try to establish the personal characteristics of a speaker in order to link other music scores to the same writer. Characteristics are, for example, the way in which clefs are drawn or the orientation of note stems. Other information about the place and time of creation can be derived from watermarks in the paper or from the kind of paper itself but the identification of such information is not part of the work presented here.

So far, the task of identifying characteristics of a writer and of attributing a music score to a certain writer has been a slow and manual process. Collections found in archives often contain 1,000s of music scores and usually information about the composer and the way of distribution is only available for a small number of these. Establishing the missing information manually is often beyond the available means and also risks further deterioration of the scores. As many libraries are in the process of digitising these scores and making them available online, so that they do not need to be touched frequently by hand anymore, an excellent opportunity arises to (partly) automate the work of musicologists. The aim of the eNoteHistory project is to develop ways of characterising a writer's handwriting and to build a system for musicologists that aids their task by automating the extraction of relevant music symbols, determining the characteristics in these symbols, and then comparing them with other music scores.

After a brief literature review in Section 2, we first outline the overall system design consisting of a database of digital copies of music scores and attached sets of handwriting characteristics, as well as image processing algorithms to extract these (Section 3). Next, we give some details on the image material used in this project in Section 4. The hierarchical structure of the image analysis system is then detailed in Section 5. Finally, the conclusions and the outlook on future work are given in Section 6.

2 Related Work

With the advances in computer technology in recent years, optical character recognition (OCR) systems (e.g. Fine-Reader) have become available that take a digitised document and recognise text characters by applying image processing techniques. These systems use dictionaries and statistics of word occurences in a similar fashion to automatic speech recognition systems to overcome ambiguities after the optical recognition. While OCR of printed documents delivers recognition rates acceptable in practice, OCR of handwritten documents still presents a challenge.
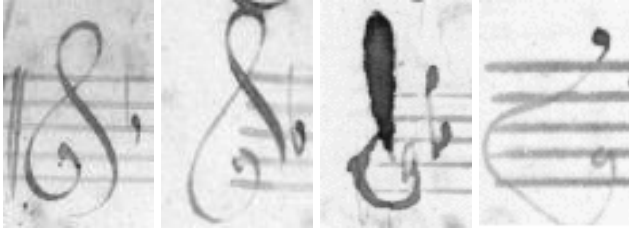

Figure 1. Treble clef drawings vary from writer to writer.

In recent years, optical music recognition (OMR) systems have also become available. Systems are either pure OMR systems (e.g. MIDISCAN, Capella-scan) or part of a larger music editing system (e.g. Finale). Common to all these systems is that they are designed to recognise printed sheet music scores where the contrast between music symbols and the background is good, staff lines are straight, and symbols are printed clearly. However, they fail at the task of recognising historic handwritten music scores because of degraded paper and ink, bent staff lines, notations that vary from writer to writer (Figure 1), symbols shining through from the reverse page, and so on [1].

Blostein and Baird [2] offer a comprehensive review of early work in this field. This includes work on forms of preprocessing to separate the music symbols from the background and noise, symbol classification, and a comprehensive overview on locating and removing staff lines (mostly for printed music scores). Armand [3] also looks at OMR for handwritten music scores but for contemporary pieces of music, not historic music scores. While the image quality of the music symbols appears to be comparable to that in our digitised scores, contemporary music scores are written on clean sheets with printed, straight staff lines. He suggests a hierarchical and recursive approach. Caldas Pinto *et al.* [1, 4] apply OMR to historic handwritten music scores with good results. They mainly use projection methods to locate music symbols. A formal approach to OMR using a grammar is found in [5].

Typically, OMR has been used to recognise music scores so that they could be played back or reprinted. Our goal of developing an OMR system to identify the writer (composer or copyist) is a new application for OMR.

3 Overall System Outline

The main goal of the final system is to offer a tool with which musicologists can easily determine the characteristics in the handwriting of a particular writer of a music score and compare it to other music scores. To do so, requires a database of music scores which contains not only a digital copy of the music score but also information about handwriting characteristics already determined. Initially, this information can be added manually into the database and thus offer a way to compare the accuracy of the automatic feature extraction with the manual process. The final system is intended to be used in two ways. Firstly, musicologists and other interested parties can search the database via the internet to find music scores with similar handwriting characteristics to their local score at hand in some archive. This way, new information about distribution paths can be collected. Secondly, the database administrators can add new records to the database when it has become clear that a particular score was written by a particular writer.

There are two ways of searching the database. Firstly, the user has a digitised copy of the score at hand and submits it to the system where a (largely) automated process is performed to determine the handwriting characteristics on that page of the score. This process is described in Section 5. The user is prompted at times to confirm or dismiss results of individual stages, so as to improve the accuracy of the system. In addition, it is possible in large music scores that more than one person has written on a page and in this case, it might be necessary for the user to select a certain part of the image as belonging to one writer. Since a music score usually consists of more than one page, the characteristics determined on various pages are collected and averaged to give an overall set of characteristics for that score (or one set for each writer if more than one person has written it). This set of characteristics can then be compared with other sets stored in the database to find music scores with similar handwriting characteristics. The output of such a query is a list of matched sets together with a similarity score for each set.

Secondly, the user can manually determine the characteristics of the handwriting in the local copy by interactively answering a number of questions about the shape and other characteristics of the present notation symbols (clefs, rests, notes, etc). Feature trees for each music symbol have been built by the musicologists in this project (based on work in [6]). The user determines the particular characteristics by moving through these trees from the tree root (representing the most general level) to the tree leaves (representing the most detailed description of the symbol). Each hierarchy level is illustrated by an idealised pictorial representation, rather than depending on a textual description which was found to be more ambiguous. The result of this manual way is another set of characteristics which can be

Figure 2. Digitised music scores: A well-preserved score (left), a stained score (centre), and a score with notation symbols shining through from the reverse page (right).

compared to other sets in the database.

Despite both ways resulting in sets of characteristics, it is important to note that the two sets do not entirely contain the same elements. There are several reasons for that. Firstly, the automated way relies more heavily on metric measures such as the average inclination of note stems, for example, than the manual way. The latter, through the feature trees, represents more a way of conceptualising the way of drawing a symbol, rather than measuring it. Such information is easy to understand for a user but hard to implement in an automated system. Secondly, the set of characteristics determined manually is linked to a particular writer, whereas the set of characteristics determined automatically is specific for a particular music score. Obviously, there are links between these two but they are not the same because a writer will typically have written more than one score.

4   The Image Material

In this project, we use a selection of music scores from the collection of the Prince Friedrich Ludwig of Wuerttemberg-Stuttgart and the Duchess Luise-Friederike of Mecklenburg-Schwerin, held at the library of the University of Rostock, Germany, as an example. This collection contains more than 10,000 music scores from the 17th and 18th century, with a large set of Wuerttembergiana from the first half of the 18th century forming an important part of the collection. Some parts of the collection have been characterised and identified by hand [6]. That work forms the theoretical basis for the manual way of determining the handwriting characteristics as discussed above.

In the first part of the current project, a selection of about 100 known and unknown writers with a total amount of 1000 pages of music scores is taken from the collection, digitised using a ProServ Dual Profi+ scanner and used as a base for testing algorithms. The music scores are digitised at 300dpi with 24bit colour information and stored in a lossless TIFF format. Figure 2 shows examples of digi-

tised scores. The left image shows a well-preserved music score, while the image in the centre presents a deteriorated score with stains. Generally, the paper has often turned yellow. It is of inhomogenous texture and stains can be found. Sometimes music symbols from the reverse page shine through and have a similar intensity and colour as the music symbols on the front page (right image).

5   Outline of Automated Handwriting Identification System

We will now focus on the automated way of determining handwriting characteristics. The system is designed in a hierarchical, bottom-up way with four levels (Figure 3). In the first level, the digitised scores undergo a preprocessing which consists of smoothing, histogram equalisation, segmentation, and morphological operations. In the second level, the preprocessed images are analysed by splitting the segmented foreground information into five layers. The layers contain primitives that may form music notation symbols. In the third level, the information from the layers is used to recognise music notation symbols of interest. Information about the objects is collected in a set of characteristics. Finally, a classification of the given score page is carried out based on the information in the set.

Information travels in both directions between the levels because it is often necessary to refine the results of a previous level after processing in the current level. The image analysis and object recognition levels also use consistency checks to eliminate irrelevant structures from the list of candidates of music symbols.

It should be noted that to find characteristics of a writer's handwriting and identify the writer, it is not necessary to select all notation symbols on a page. Badly segmented or superimposing symbols can be left out which is different from OMR for automatic recovery of music from a music score where all symbols must be identified.

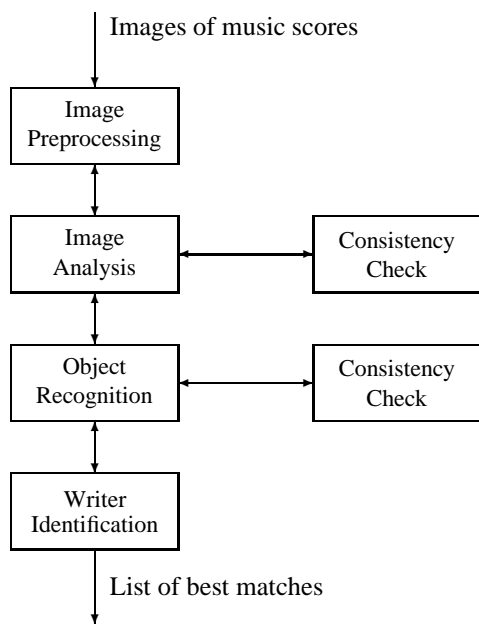In the following, we give details about each of the

Figure 3. Schematic outline of the hierarchical system.

four levels of the hierarchical system. Results of some of the various operations performed are shown in Figures 4, 5 and 6.

5.1 Image Preprocessing

The aim of the preprocessing is to prepare the way for the following image analysis, so as to reduce the number of errors in that stage. First of all, the image is smoothed with a Gaussian filter to reduce the amount of image noise (Figure 4). In a second step, histogram equalisation is performed separately on each colour channel. Ideally, there would be two distinct, separate peaks for the foreground and background. As can be seen from the histogram of the blue channel before equalisation (Figure 5, right), this is not the case in reality. The intensities of foreground and background pixels overlap. The crossed area denotes the intensities related to pixels of music symbols. The left and centre images show the effect of histogram equalisation on the blue channel. The contrast becomes much better and it is thus easier to segment the image into foreground and background areas (see below).

Next, by rotating the image at small angular steps and applying a horizontal projection at each step, we find the angle that maximises the values for the horizontal projection. The image is rotated by this angle to counter errors introduced at the time of digitising the music sheet.

As we can see in the histogram (Figure 5, right), there is considerable overlap in the intensities of foreground and background if the image is taken as a whole. Applying a threshold on a global level for the whole image would
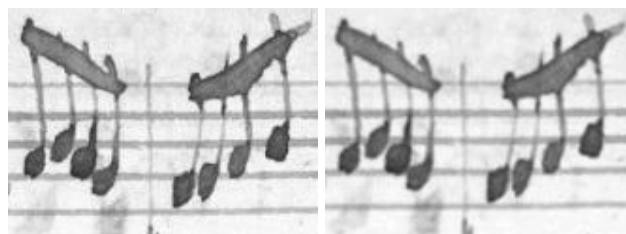

Figure 4. Reducing image noise by smoothing with a Gaussian filter: before (left) and after (right).

therefore not lead to good segmentation results. By tiling the image into small areas of size 30×30 pixels and performing local thresholding on each of these tiles, better results are achieved compared to a global method. Thresholding is perfomed using Otsu's method [7].

Finally, the morphological operations of erosion and dilation are performed to remove "salt and pepper" noise. Elements smaller than 4 pixels in diameter are removed this way and the images become clearer. In addition, the areas to the left and right of the staff lines are also deleted (or simply marked as not containing information for the analysis stage) to reduce errors. Similarly, after the staves have been detected (see below), areas at the top and the bottom of a page outside the stave area can be deleted as well.

5.2 Image Analysis

The goal of the image analysis level is to decompose the segmented foreground pixels into primitives of five layers:

1. horizontal lines as candidates for staff lines,

2. vertical lines as candidates for bar lines and note stems,

3. small round objects as candidates for note heads,

4. polyline structures as candidates for clefs etc,

5. other information, e.g. text.

Finding staff lines is done in a two-stage process. First, we look for the staves on a block level without determining the position of each individual line (Figure 6, centre) by applying a combination of an edge detection (Sobel operator) and horizontal projection method. Peaks in the horizontal projection correspond to areas of background pixels, valleys to the foreground pixels and in particular the staff lines. Small distances between valleys correspond to the staff spacing, while larger distances relate to the areas between staves. These larger distances are determined in a dynamic thesholding operation on the horizontal projection.

Then, within each staff, the individual staff lines are determined. To do so, we take the image after edge detection and let 20 equally spaced "search rays" run vertically through the staves. Each time an edge is hit by a
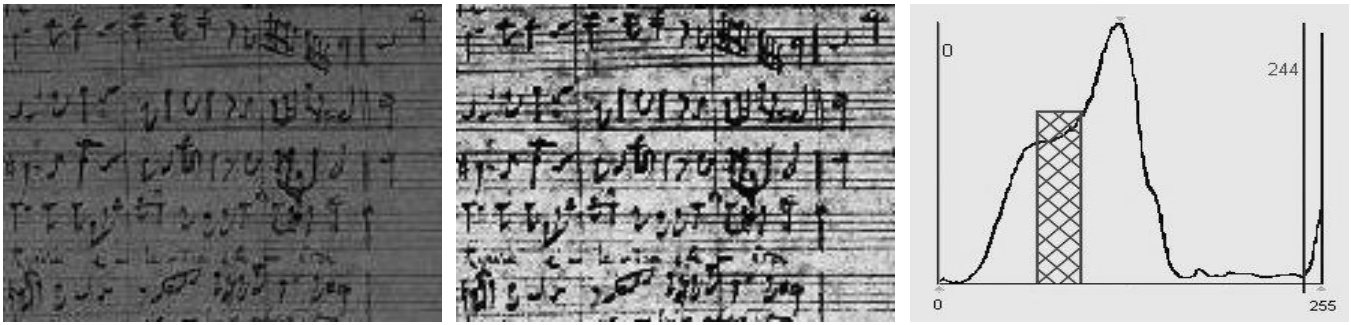
Figure 5. Histogram equalisation: before (left) and after (centre). Right: Histogram of the blue channel before histogram equalisation (crossed area marks intensities of music notation symbols).

search ray, the vertical position is noted. A histogram of the occurrence of distances from one position on a ray to the next position is computed and the average thickness of a staff line as well as the average distance between two staff lines in a staff are determined. Next, a template is created with five horizontal lines of average thickness and with average distances between them. The width of the template is 25 pixels. This template is moved across each staff, the normalized cross correlation is calculated at each position and the best match is taken as the starting point for exactly following the lines. This is also done by template matching but at each horizontal position, the template is only moved vertically to find the best matching position (Figure 6, right). Since the segmentation process can accidentally remove parts of staff lines at times, the segments found at each horizontal position are joined and missing staff lines are added.

The staves serve as a reference system for the other layers. Bar lines and note stems are found by vertical projection within each system of five staff lines. The length distinguishes bar lines from note stems. Bar lines run at least from the first staff line to the fifth and last staff line, but may actually cross these. Note stems are shorter and must also be attached to a note head. Furthermore, the spacing between two bar lines is much larger than between two consecutive note stems which adds additional certainty to the process.

Candidates for the note heads are found by a closing and opening operation with a circle as structure element. Only circular shapes that have a diameter similar to the distance between two staff lines are kept. We currently compare these results with a template matching search where a template is created from a circle with a diameter equal to the distance between two staff lines. Positions where the normalized cross correlation value is ≥ 0.8 are kept. In either case, the candidate positions are checked for consistency with the positions of the note stems. Stem candidates without an associated note head are removed from the list of candidates. Note heads can occur without a corresponding note stem (whole and half notes) but note stems can

never occur without a corresponding note head.

Template matching is currently used to find complex, polyline symbols such as clefs, accidentals, flags etc. The templates are generic templates designed by the musicologists in this project. They are sufficient to identify the rough positions of the symbols. Once these are known, the areas are characterised by a principal component analysis (PCA). Such an approach has been used in other application areas, such as visual speech recognition systems [8]. Tests are currently performed to see if this is a viable way. Any remaining image structures are put into the last layer.

5.3   Object Recognition and Writer Identification

These two levels are currently being implemented and tested. The object recognition level uses the structural knowledge we have about music scores and tries to find corresponding primitives in the five layers. The staves serve as a reference system. Music notation symbols significantly outside the staves are disregarded. As already mentioned, it is not a requirement to find all notation symbols in an image, as long as a sufficient number is found to determine the handwriting characteristics of the writer. An object, for example a music note, is formed by the various (possible) primitives, such as a note head, stem and flag. The lists of candidate positions for note heads, stems and flags are parsed simultaneously to find corresponding primitives which are then stored as joint objects.

Characteristics that are currently determined are the position of the note stem relative to the note head (left, centre, right; separately for upwards and downwards facing note stems), the length and inclination of note stems, the distance between two staff lines, and the way complex notation symbols such as clefs, flags and rests are drawn. Complex symbols are characterised by their principal components. Usually, the first five principal components are sufficient to capture more than 90% of the variance.

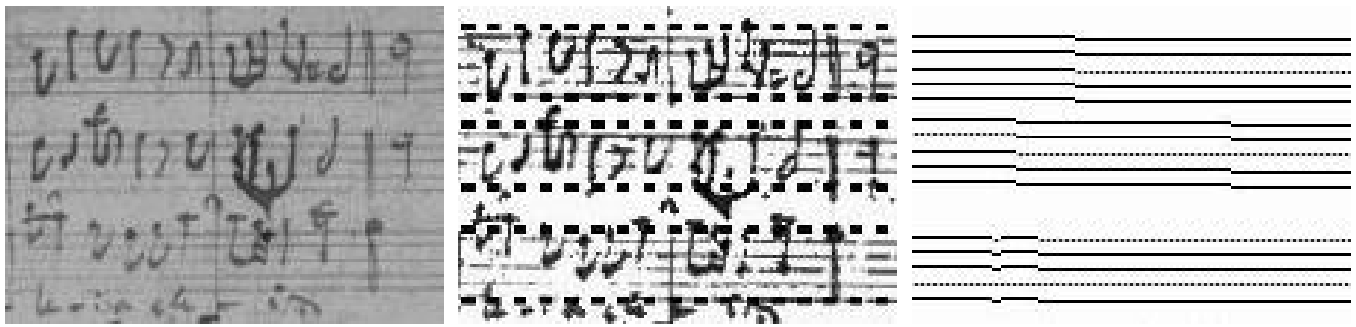Writer identification is performed in two ways due to

Figure 6. Original intensity image (left). Dashed lines denote detected staves after preprocessing (centre). Staff line detection (right).

the differences between positional and metric information. Positional information such as that of note heads and stems can be compared by simply comparing the categories the information falls into. For metric information, a statistical similarity measure is used. Several similarity measures are considered and tested for optimal results. A simple similarity measure is, for example, the sum of absolute differences. The result of the identification process is a list of similarity scores for the match of the current music score with other music scores in the database.

6 Conclusions and Further Work

We have presented the current status of this ongoing project on developing an OMR system for writer identification to aid the work of musicologists. The final system consists of a database of digitised music scores and methods to determine the handwriting characteristics of a writer. This can be done in an interactive process using feature trees as well as in an automated process using image processing techniques. The latter has been presented here. It is expected that the system will not only speed up the process of identifying a writer but also give the musicologists new insight into the kind of features that can be used for identification.

The system we currently develop is a hierarchical system. The first two of the four levels have been implemented, while the remaining two levels - the object recognition and writer identification - are under development. The algorithms need to be tested further before they are integrated in the overall system together with the database and its searching facilities. We also want to extend the set of characteristics to include accidentals and chords.

7 Acknowledgement

References

[1] P. Vieira and J. Caldas Pinto, "Recognition of Musical Symbols in Ancient Manuscripts," in *Proc 2001 Int Conf on Image Processing ICIP-2001*, Thessaloniki, Greece, Oct. 2001, vol. 3, pp. 38–41, IEEE.

[2] D. Blostein and H.S. Baird, "A Critical Survey of Music Image Analysis," in *Structured Document Image Analysis*, H. Baird, H. Banke, and K. Yamamoto, Eds., Berlin, 1992, pp. 405–434, Springer.

[3] J.-P. Armand, "Musical Score Recognition: A Hierarchical and Recursive Approach," in *Proc 2nd Int Conf on Document Analysis and Recognition ICDAR'93*, Tsukuba, Japan, Oct. 1993, pp. 906–909.

[4] J. Caldas Pinto, P. Vieira, M. Ramalho, M. Mengucci, P. Pina, and F. Muge, "Ancient Music Recovery for Digital Libraries," in *Research and Advanced Technology for Digital Libraries*, J. Borbinha and T. Baker, Eds., Berlin, Germany, 2000, pp. 24–34, Springer.

[5] B. Coüasnon and J. Camillerapp, "A Way to Separate Knowledge from Program in Structured Document Analysis: Application to Optical Music Recognition," in *Proceedings of the 3rd International Conference on Document Analysis and Recognition ICDAR'95*, Montreal, Canada, Aug. 1995, vol. 2, pp. 1092–1097.

[6] E. Krüger, *Die Musikaliensammlungen des Erbprinzen Friedrich Ludwig von Württemberg-Stuttgart und der Herzogin Luise-Friederike von Mecklenburg-Schwerin in der Universittsbibliothek Rostock*, Ph.D. thesis, University of Rostock, Germany, 2003, submitted.

[7] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[8] U. Meier, R. Stiefelhagen, J. Yang, and A. Waibel, "Towards Unrestricted Lip Reading," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 14, no. 5, pp. 571–585, 2000.