A ROBUST DETECTOR FOR MUSIC STAVES

I. LEPLUMEY */ J. CAMILLERAPP */ G. LORETTE **

*/ INSA - 20, Avenue des Buttes de Cœsmes 35043 RENNES CEDEX - email: Ivan.Leplumey@irisa.fr
** IFSIC - Campus de Beaulieu Avenue du Général Leclerc 35042 RENNES CEDEX - email: Guy.Lorette@irisa.fr

Abstract
The context of this paper is the automatic recognition of music staves. We present a method, based on a prediction-and-check technique, to extract staves. It can detect lines with some curvature, discontinuities and inclination. Lines are asserted to be part of a staff if they can be grouped by five, thus completing the staff. This last phase also identifies additional staff lines.

1. Introduction
The context of this paper is the automatic recognition of binary images of staves. Few papers have been published about this problem. A summary is in [1]. A score is viewed as a multi-layer document. The lowest-level layer corresponds to staves, the next level corresponds to measures. The third level corresponds to the notes (pitch, duration) and the final level contains all the textual annotations for performance.

An important step, stated in all papers, is the recognition of staff lines. Some methods [3] use point counts on each line of the image. But these projective techniques do not easily allow the detection of staff lines having an inclination; furthermore they are sensitive to the curvature of the lines.

Similarly, the method of labelling white components of an image, which we developed previously to extract staff lines, is highly sensitive to discontinuities [5] [6]. In order to improve its results, one needs to revise the initial segmentation of white areas.

In [7] the global inclination of scores is found by the string method, which finds the longest straight line for each point of the central column of the image. This global view is unable to deal with some curvature in staff lines.

The methods above do not have the same sensitivity to the characteristics mentioned above: staff line slope, curvature and discontinuities. The method developed by N.P. Carter [2] seems to be the most robust with regard to these factors; it is based on an analysis of a Line Adjacency Graph [8].

We wish to develop a method which is robust with regard to the scale factor, the inclination of staff lines, a light curvature of these lines and discontinuities in them. These noises can stem from the quality of the document or the digitization technology.

We first present the extraction of the scale factor, then the following and extraction of lines, and finally the recognition of staves.

2. Thickness of staff lines and interlines
The initial binary image, a matrix, is first mapped, column by column in a sequence of black areas identified by their ordinate of their origin and end. Such an area we call a run length. This mapping, usual in image compression, also allows a speed improvement in later steps of the algorithm.

We assume, as in [4], that the maximum of the histogram of run lengths' thickness is a good estimate of the thickness of staff lines along the vertical axis. This value will be called T_black (fig. 1). Similarly, the maximum of the histogram of distances between two consecutive run lengths of a column is a good estimate of interline thickness. We call it T_white (fig. 1).

![Figure 1. Thickness parameters](image)

From here, we call observation a run length. To detect staff lines, we define compatible observations. They are run lengths whose length is T_black +/- 2 pixels.
This allowance is justified by the sampling error on the ends of run lengths, which is +/- 1 pixel.

3. Lines detection by a two-level prediction-and-check strategy

3.1. Overview

Starting from compatible observations, we build some hypotheses on the presence of lines by grouping compatible observations into lines. But these groups cannot be directly interpreted as staff lines because some musical objects (e.g. links) have a similar local definition. In order to conclude decisively, we need a larger context. Groupings are built by two levels of prediction-and-check:

- the first level tries to accumulate \( n_0 \) compatible observations, in order to ensure we are in the presence of the beginning of a line.
- from a beginning of line, the second level tries to maximize the line length.

Groups are named stains in the first stage and dashes in the second.

At the first level, we have a very rough knowledge of the inclination, which explains the highly loose tests for grouping. This forbids discontinuities. On the contrary, in the second level, the knowledge of the inclination is sufficient enough that we can allow a momentary loss of observations.

**Figure 2. Stains and dashes**

For now, a compatible observation can be considered only once. We first try to integrate it in a dash. If this fails, we try to group it with a stain. Else, we create a new stain. We now review each level.

3.2. Stain level

**Creation:** a compatible observation not grouped with a stain or a dash generates a new stain.

**Prediction:** for a stain followed until column \( N \), the prediction of the observation at column \( N+1 \) is a run length with the same ordinates as the observation associated to the stain at column \( N \), and with abscissa \( N+1 \).

**Verification:** an observation matches the prediction of a stain if it is compatible (thickness criterion) and if the intersection between the observation and the prediction is not empty (position criterion). A stain is a 4-connected component of compatible observations.

**Releasing:** the absence of an observation matching a prediction leads to the destruction of the stain.

3.3. Dash level

**Creation:** a stain accumulating \( n_0 \) observations creates a dash. We link this number to the scale of the document. A too low value creates too many dashes, while a too high one, by delaying the detection of dashes, impeaches a complete detection. We therefore take as threshold \( n_0 \) the value of \( T_{\text{white}} \) defined in 2.

**Prediction:** here the prediction is finer than the one at the stain level. Indeed, since we have enough observations, and assuming that dashes are locally equivalent to straight segments, we can find the least squares line by storing, for each dash, the parameters \( \Sigma x, \Sigma y, \Sigma xy, \Sigma x^2 \) where \( x, y \) are the ordinates of the centers of the observations. From these values, it is easy to predict center of the run length in the current column.

**Verification:** an observation fulfills the prediction associated to a dash if it is compatible (thickness criterion) and if the distance between the predicted center and the observed one is at most 2 pixels (position criterion). The four parameters \( \Sigma x, \Sigma y, \Sigma xy, \Sigma x^2 \) used for the prediction are then locally adjusted in order to allow a light curvature of staff lines, by this formula:

\[
\text{Parameter}(N+1) = (n_0 - 1) * \text{Parameter}(N) / n_0 + (\text{Corresponding value in the observation})
\]

We thus weight-balance the past and the present.

If no correct observation is present (white areas, too wide run length or incorrectly centered run with respect to the prediction), the parameters keep their previous values.

**Halting:** if, for \( n_1 \) consecutive columns, center predictions cover white points of the initial image, that is if no run length is present at the predicted position, then the dash is not lengthened. It is stored for later interpretation. This strategy allows for discontinuities in known dashes. The choice of \( T_{\text{white}} \) for the thickness \( n_1 \) of discontinuities seems a good compromise. It is not at all a critical value.

3.4. Conclusion

The technique described above fulfills the stated requirements and is a robust detector of lines. The use of a local predicting function for the dash level prevents an
excessive fragmentation of staff lines, thus reducing the cost of the interpretation phase. The method cost is linear in the number of columns, of run lengths per column, of stains and of dashes.

4. Stave detection

Dashes obtained in the previous phase can have multiple interpretations: staff lines, additional lines, links, parts of texts. So we need to extract from these dashes those which correspond to staff lines and to reunite the fragmented staff lines.

4.1. Construction of an interpretation graph

A staff is defined as a set of five parallel and equidistant lines. The distance between to consecutive staff lines is \( d_4 = T_{black} + T_{white} \). We first remove dashes which have too much angular variation between the beginning and the end. This eliminates dashes with too much curvature (which is the case for some links). The allowed difference is arbitrarily set to 10 degrees. A dash is now represented as a segment joining the creation and termination points. To compare two dashes, we use the symbols in figure 3. We represent by a graph the knowledge about the image. Each dash becomes a vertex labelled by its length.

![Figure 3. Comparison interval of the positions of two dashes](image)

Vertices (dashes) \( N_1 \) and \( N_2 \) are connected if their abscissa-comparison interval is not empty and if and only if:

\[
\begin{align*}
(d_4 - \varepsilon) & \leq y_A - y_I \leq (d_4 + \varepsilon) \\
(d_4 - \varepsilon) & \leq y_I - y_D \leq (d_4 + \varepsilon)
\end{align*}
\]

where \( y_p \) is the ordinate of point \( p \).

The allowance \( \varepsilon \) is equal to twice the uncertainty on run lengths thickness. We thus check the position and parallelism of lines.

We now need to check if each connected component of the graph can be interpreted as a staff.

4.2. Interpretation of graph vertices

Each connected component is rewritten: if the vertex has different predecessors (resp. successors) they are merged into one vertex labelled with the sum of their values. Outward (resp. inward) edges of the predecessors are also merged into one edge. If a connected component has less than five vertices, it does not correspond to a staff. If it has five (resp. more) vertices it corresponds to a staff without (resp. with) additional lines. The next step is to find where in the connected component are the main lines of the staff. This is done by searching in the component a sequence of five consecutive vertices for which the sum of the labels is maximum, which is equivalent to the identification of lines of maximum size. We thus obtain the five main lines, and the interpretation of the additional ones.

5. Results

The example image is a 300 dpi scanned image with a 5 degree inclination and with additional lines (fig. 5). Its width is 2600 columns, each with an average of 60 run lengths. The number of stains per column is low, approximately 5 stains per column. Fig. 4 shows the history of the number of dashes. The ascending curve is the number of ended dashes respective to the number of the column. The curve with a plateau corresponds to currently being lengthening dashes. The plateau’s height (roughly 50) is slightly more than the number of staff lines in the image. The descending final part is explained by the inclination of the score.

At the end of the lines extraction step, 150 lines remain vs only 50 staff lines, and about 15 additional ones. Only one staff line remained broken in two parts. Finally, after the staff-line interpretation step (fig. 6), we can observe the different staves identified by a number, with a thin drawing for their five main lines and a thick one for the additional lines. The method has been tested on ten images with inclinations from 0 to 10 degrees and a grain of 300 or 400 dpi. In all images we correctly identify staves and their constituents. Additional lines not covered by musical symbols are generally identified, although this was not our original goal.

![Figure 4. Number of ended or not dashes for each column of the image](image)
6. Conclusion

This method for staves detection allows, thru a two-stage prediction-and-check strategy, a robust staff detection against the scale factor and inclination. It also handles possible defects in staff lines (curvature, discontinuities). This robustness has been verified on images with differing characteristics. Ongoing work includes a generalization of the lines detector with a Kalman Filter, able to detect all linear musical objects, that is bars, notes, quavers and even links. This leads to slightly modifying the line-detector’s parameters and, notably, to ignoring the estimate thickness of staff lines.

Acknowledgments

This work has been done in close cooperation with Mr. Bernard RETIF, professor of Musical Computer Science at the University Rennes II

References

[1] D. BLOSTEIN, H. S. BAIRD
"A Critical Survey of Music Image Analysis”
Structured Document Image Analysis, Springer Verlag 1992, pp. 405-434

"Automatic Recognition of Printed Music in the Context of Electronic Publishing”
PhD thesis, Univ. of Surrey, Depts. of Physics and Music, 1989

[3] I. FUJINAGA
"Optical music recognition using projections”
Master’s thesis, McGill University, Faculty of Music, Montréal, CANADA, 1988

"The recognition system of printed piano music”
Structured Document Image Analysis, Springer Verlag 1992, pp. 435-454

"Comparison of region labelling for musical scores”
ICDAR - IAPR Conference on Document Analysis and Recognition, Saint-Malo, 1992, pp. 674-682

"Coopération entre la segmentation des régions blanches et des régions noires pour l’analyse des partitions musicales”
Séminaire de Recherche en Reconnaissances des Formes et Intelligence Artificielle (RFIA AFCET)
Lyon Villeurbanne, 1991, pp. 1045-1052

[7] P. MARTIN
"Réseaux de Neurones Artificiels: Application à la Reconnaissance Optique de Partitions Musicales”
Thèse, Univ. JOSEPH FOURIER, GRENOBLE I, Laboratoire TIM 3, Institut IMAG, 1992

[8] T. PAVLIDIS
"Algorithms for graphics and image processing” Springer-Verlag, Berlin-Heidelberg, 1982