Content aware music score pre-processing

Ana Rebelo and Jaime S. Cardoso
INESC Porto, Faculdade de Engenharia, Universidade do Porto, Portugal
{arebelo, jaime.cardoso}@inescporto.pt

Abstract

In most OMR algorithms, the pre-processing stage consists on a binarization operation followed by the estimation of the staffline thickness and the vertical line distance within the same staff. Subsequent operation can use these values as references. This work presents an improvement on previous conventional estimates for the two reference lengths and consequently a novel binarization method based in the content knowledge of the image. An objective evaluation over a large dataset of handwritten music scores is conducted based in the performance of the subsequent steps in an OMR system.

1. Introduction

Printed documents and handwritten manuscripts deteriorate over time, causing a significant amount of information to be permanently lost. Among such perishable documents, musical scores are especially problematic. Digitization has been commonly used as a possible tool for preservation, offering easy duplications, distribution, and digital processing. However, to transform the paper-based music scores and manuscripts into a machine-readable symbolic format an Optical Music Recognition (OMR) system is needed. Most OMR algorithms rely on an estimation of the staffline thickness and the vertical line distance within the same staff after an image binarization operation.

2. Robust estimation of staff line thickness and spacing in the gray-scale domain

The conventional estimation of the staffline thickness and spacing assumes the run-length encoding (RLE) of each column of the *binary* music score. In this representation, the most common black-run is likely to represent the staffline thickness (staffline_height) and the most common white-run is likely to represent the staffline spacing (staffspace_height) [3]. Although the performance of this conventional method is excellent in printed music scores and very good in handwritten scores, the estimation fails under severe degradation of the scores. We suggest a more robustly estimation of the sum of staffline_height and staffspace_height, hereafter termed `line_thickness+spacing`, by finding the most common sum of two consecutive vertical runs (either black run followed by white run or the reverse). The process is illustrated in Figure 1.

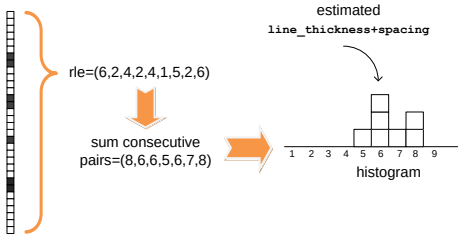

Figure 1. Illustration of the estimation of the reference value `line_thickness+spacing` using a single column.

The mean value of errors for the reference value `line_thickness+spacing` using a dataset of 50 binarized handwritten music scores with Otsu's method and the references manually measured by three independent individuals are: for the conventional estimation 2.4px and for the proposed estimation 0.4px. We also proffered to compute the histogram of the runs for 'every' possible binary image, by accumulating the runs' frequency when varying the binarization threshold from the lowest to the highest possible values. This procedure of computing the reference length `line_thickness+spacing` without assuming any binarization threshold, allows the extraction of important information directly from the gray-scale image (mean value of error: 0.4px).

3. Content aware music score binarization

A binarization method designed to maximize the number of the pairs of consecutive runs summing

line_thickness+spacing will likely maximize the quality of the binarized lines, improving the following operations in an OMR system. However, the direct maximization could lead to a threshold value resulting in a destructive quantity of noise. The use of relative histograms is also prone to problems since now one may end up choosing a threshold with a very low absolute count of runs in line_thickness+spacing. Therefore, we restrict the candidate thresholds to those producing a histogram of runs with the mode at line_thickness+spacing. If no threshold is found we consider the minimum integer i for which there are threshold values with histogram mode at line_thickness+spacing $\pm i$. From the set of candidate thresholds, the proposed binarization method for music scores simply selects the threshold that maximizes the count of pairs of consecutive runs on the mode.

This global thresholding method can also be applied to a sampling window around a pixel p, converting the proposed method to a local method. We suggest computing a single threshold per column, using as window a vertical strip with height equal to the height of the image and width defined by the user. In order to reduce the computational cost, the threshold value is computed on a set of sampled columns, and then, for the rest of the columns, the threshold value is calculated by interpolating on the sampled columns.

4. Experimental evaluation

A dataset of music scores composed of 65 handwritten scores, from 6 different authors was used for experimental validation. All the scores in the dataset were reduced to gray level information.

The methods chosen try to encompass different categories of thresholding operations: Kapur, Sahoo and Tsallis are entropy methods, Huang is based on object attributes, Otsu is a clustering method, Tsai's moment preserving and Khashman's luminance are histogram based methods.

Some of the algorithms tested required the input of different parameters that were obtained by experimental testing.

For global thresholding processes, three different approaches were taken for evaluation: comparison of the resulting thresholds with an optimal supervised threshold for each image; Misclassification Error (ME), by comparing resulting images pixel by pixel to ground truth binarizations; comparison between results of staff finder algorithms applied to each binarized image (Stable Path [1] and Dalitz [2] methods).

The Binarization based in LIne Spacing and Thickness (BLIST) method, proposed in this article, performed above average. Entropy based binarizations and Khashman's algorithms got fairly similar results to each other. Huang and Tsai managed to top these results, with acceptable line detection rates (2.6% and 2.1%, respectively) and ME (6.2% and 4.7%, respectively). There are, however, two binarization techniques that get consistently better results than the others: Otsu and BLIST methods. The only major difference is the higher missed staff detection rate for the Otsu's algorithm (18.8%).

Global methods can generally produce good outputs. Even so, for some of the scores there is not possible to find a single threshold value. The window width for the adaptive version of the BLIST method used was a fixed percentage of the total image width. The interpolation of the threshold values obtained was generated with a third degree polynomial regression. Otsu's method, having good results among global methods was also implemented as adaptive, using the same reasoning. The input parameters were obtained experimentally.

When testing local thresholding methods, two new error rates are presented: the Missed Object Pixel (MOPx) rate and the False Object Pixel (FOPx), dealing with loss in object pixels and excess noise, respectively.

Ad BLIST (15.6%) and YB (12.4%) show the lowest MOPx, meaning these are the methods that find most of the correct pixels, which translates into lower missed staves rates. Even so, Ad BLIST also has a FOPx rate slightly higher than the other methods (18.5%). This higher noise also translates into a slightly higher false staves rate with Dalitz method (3.8%). Bernsen's binarizations, although presenting the highest missed pixel rate (9.0%), seem to perform well in the staff finding steps, having both the lowest missed (1.3%) and false (1.9%) staves rates.

5. Conclusion

We believe the results of the adaptive BLIST method are affected by the left and right margins of the scores, because they do not carry significant information for the algorithm. In this manner one possible way to guide this work is to somehow treat this issue. Another way is to use global BLIST for the staff finding process and a new method for the next OMR steps.

References

[1] J. S. Cardoso, A. Capela, A. Rebelo, C. Guedes, and J. F. P. da Costa. Staff detection with stable paths. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(6):1134–1139, 2009.

[2] C. Dalitz, M. Droettboom, B. Czerwinski, and I. Fujigana. A comparative study of staff removal algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:753–766, 2008.

[3] I. Fujinaga. Staff detection and removal. In S. George, editor, *Visual Perception of Music Notation: On-Line and Off-Line Recognition*, pages 1–39. Idea Group Inc., 2004.