Robust staffline thickness and distance estimation in binary and gray-level music scores

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

***Abstract*—The optical recognition of handwritten musical scores by computers remains far from ideal. Most OMR algorithms rely on an estimation of the staffline thickness and the vertical line distance within the same staff. Subsequent operation can use these values as references, dismissing the need for some predetermined threshold values. In this work we improve on previous conventional estimates for these two reference lengths. We start by proposing a new method for binarized music scores and then extend the approach for gray-level music scores. An experimental study with 50 images is used to assess the interest of the novel method.**

***Keywords*-Optical music recognition; document image processing; image analysis**

I. 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 (facilitating operations such as search, retrieval and analysis), an Optical Music Recognition (OMR) system is needed. This justifies the research on reliable OMR algorithms [1], [2].

Most OMR algorithms rely on an estimation of the staffline thickness (staffline_height) and the vertical line distance within the same staff (staffspace_height), see Figure 1) [3], [4]. Further processing can be performed based on these values and not be dependent on some predetermined magic numbers. The use of fixed threshold numbers, as found in other areas, makes systems inflexible and difficult to adapt to new and unexpected situations. As shown by Fujinaga [5], these values can be estimated with good accuracy as the most frequent black (staffline_height) and white (staffspace_height) vertical runlength, respectively.

Nevertheless, for handwritten music scores strongly contaminated by noise – either due to the low quality of the original paper in which it is written or due to defects introduced during digitalization and binarization – the results are still unsatisfactory. This impairs the quality of subsequent operations, namely the detection and removal of stafflines.

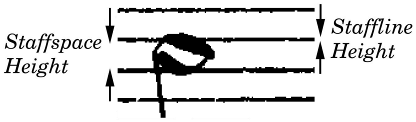


Figure 1. The characteristic page dimensions staffline_height and staffspace_height.

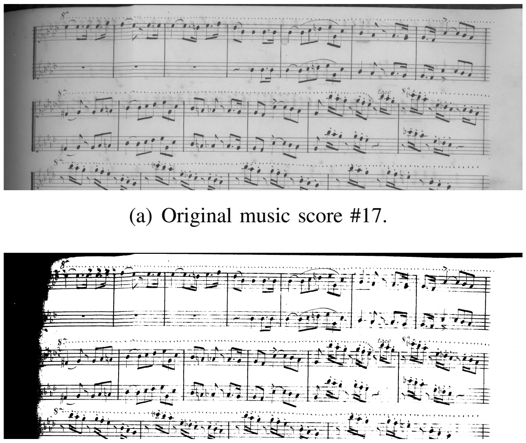
In this work we introduce two main contributions: a robust method to estimate the staffline_height and staffspace_height on binarized music scores and the generalization to gray-level music scores.

II. CONVENTIONAL ESTIMATION OF STAFFLINE THICKNESS AND DISTANCE

Run-length encoding (RLE) is a very simple form of data compression in which runs of data (that is, sequences in which the same data value occurs in consecutive data elements) are represented as a single data value and count. In a binary image, used as input for the recognition process here, there are only two values: one and zero. In such a case, the run-length coding is even more compact, because only the lengths of the runs are needed. For example, the sequence {1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1} can be coded as 8, 3, 13, 8, 2, assuming 1 starts a sequence (if a sequence starts with a 0, the length of zero would be used). By encoding each column of a digitized score using RLE, the most common black-runs represents the staffline_height and the most common white-runs represents the staffspace_height. Even in music scores with different staff sizes, there will be prominent peaks at the most frequent staffspaces. These estimates are also immune to severe rotation of the image [6], [5].

III. ROBUST ESTIMATION OF STAFFLINE THICKNESS AND SPACING

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, as illustrated in Figure 2. For this score the conventional estimation provides staffline_height = 1 and staffspace_height = 1 (the true values are staffline_height = 5 and staffspace_height = 19).



(a) Original music score #17.

(b) Score binarized with Otsu' method.

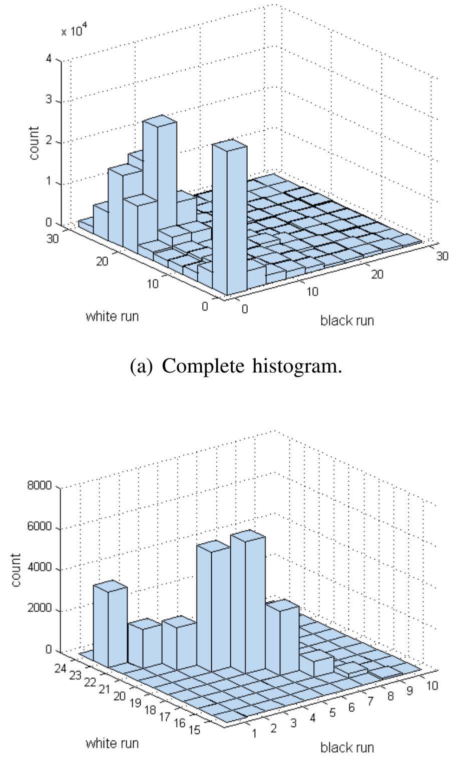
Figure 2. Unsuccessful estimation of staffline_height and staffspace_height by vertical runs.

Isolated black pixels and fluctuations in the thickness and distance between lines pose challenging difficulties. Two observations are decisive for an improved estimation of the reference lengths: first, the length of the run of white pixels before and after the isolated black pixels will vary a lot, 'randomly'. Second, a local fluctuation in the thickness of the staffline (due to noise) is often compensated by a variation with an opposite sign of the local distance between lines–this effect is visible in Figure 1.

Therefore, it seems that the estimation of the sum of staffline_height and staffspace_height can be done much more robustly by finding the most common sum of two consecutive vertical runs (either black run followed by white run or the reverse). This is illustrated in Figure 3, with the histograms of the black runs, white runs and the sum of two consecutive runs. The prominent peak at the black and white runs histograms is at run = 1, due to the noise on the binarization process. Nevertheless, the prominent peak at the sum two consecutive runs is at 24, consistent with the true values of staffline_height and staffspace_height.

We propose that, when possible, any subsequent operation should be based on this new reference length, staffline_space_height. Nevertheless, robust estimates of staffline_height and staffspace_height can now be obtained knowing that their sum should equal staffline_space_height. To reliably estimate these values we start by computing the 2D histogram of the pairs of consecutive vertical runs. Next, we select the most common pair for which the sum of the runs equals staffline_space_height.

Figure 4 shows the 2D histogram for the music score in Figure 2. As visible, restricting ourselves to the pairs summing 24 (the estimated value for the sum), we are able to recover the correct value for the line thickness and space.



(a) Complete histogram.

(b) Histogram of the pairs summing 24 (the peak value observed in Figure 3(c)).

Figure 4. Histogram of the pairs (black run, white run) for score #17 in Figure 2.

IV. Staffline thickness and spacing estimation in gray-level images

In some binarized images, the noise level is such that even the new method is unable to correctly estimate staffline_height and staffspace_height. Figure 5 shows the result for score #01. Even though the sum of staffline_height and staffspace_height is correctly estimated, the individual values are not.

A source of difficulties is the binarization algorithm itself. State-of-the-art methods fail to correctly binarize the score under conditions such as low paper quality, gradient effect on the illumination, etc. It seems that better results could be achieved directly on the gray-level score.

Instead of computing the histogram of the runs for a single binarized score, using a threshold computed by a state-of-the-art binarization method, we propose to compute the histogram of the runs for 'every' possible binary image by varying the threshold from a low to a high limit.

The rationale is that, although binarization algorithms have difficulties in finding a proper threshold, there is an interval of values that produce a proper binarized image and that will contribute to a robust histogram; threshold values outside this interval will likely produce 'random' runs that will disperse over the histogram. Figure 6 shows the histograms accumulating the runs for threshold from 1 to the median value of the image (we assume that in a music score most of the pixels are background and therefore the median pixel value is background).

Now, we are able to correctly estimate not only the sum of staffline_height and staffspace_height but also the individual values. Moreover, the histogram for the sum shows a much more prominent peak, suggesting a much more robust estimate of this length.

V. Experimental Assessment

The proposed methodology was tested on a modern set of 50 handwritten music scores. The reference staffline_height and staffspace_height were manually measured by three independent individuals. The binarized version of the scores were obtained with the Otsu threshold algorithm, as implemented in the Gamera project[1]. The performance obtained over the dataset is summarized in Table I.

We see an improvement in the estimation of both parameters when adopting the novel approach in binary images. When applying the approach for gray-level images a further improvement is achieved. We also notice that the sum of of staffline_height and staffspace_height is the most reliable estimation.

VI. Conclusion

Studies on music recognition are under way to build up music databases and automatic performance. Since staffs are important symbols to determine the position or size of other music symbols in these studies, almost all methods extract its position in initial recognition. Before the stave candidate point is extracted, the line width and interval of the staff is usually estimated to work as reference lengths for the subsequent operations.

We presented a robust method to reliably estimate the thickness of the lines and the interline distance. We assumed that the initial image is converted, column by column, in the run-length coding. Next, we introduced the estimation of the sum of the two lengths as a more reliable estimation than the independent estimation of the two lengths. The individual lengths were then estimated as the most likely combination of runs summing to the pre-estimated total. To overcome the difficulties of binarization algorithms with low quality music scores, we propose to integrate the estimation over every possible binarization threshold.

The basic idea of estimating first the sum of quantities of interest, and then estimating the individual quantities of interest with the constraint that they sum to the estimated value, may apply in other areas of document image analysis, or in general image analysis.

Acknowledgment

This work was partially funded by Fundação para a Ciência e a Tecnologia (FCT) - Portugal through project PTDC/EIA/71225/2006.

References

[1] D. Blostein and H. S. Baird, "A critical survey of music image analysis," in *Structured Document Image Analysis*, Baird, Bunke, and Y. (Eds.), Eds. Heidelberg: Springer-Verlag, 1992, pp. 405–434.

[2] H. Miyao and M. Okamoto, "Stave extraction for printed music scores using DP matching," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 8, pp. 208–215, 2004.

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

[4] J. S. Cardoso, A. Capela, A. Rebelo, C. Guedes, and J. F. P. da Costa, "Staff detection with stable paths," *IEEE Transactions Pattern Analysis Machine Intelligence*, vol. 31, pp. 1134–1139, 2009.

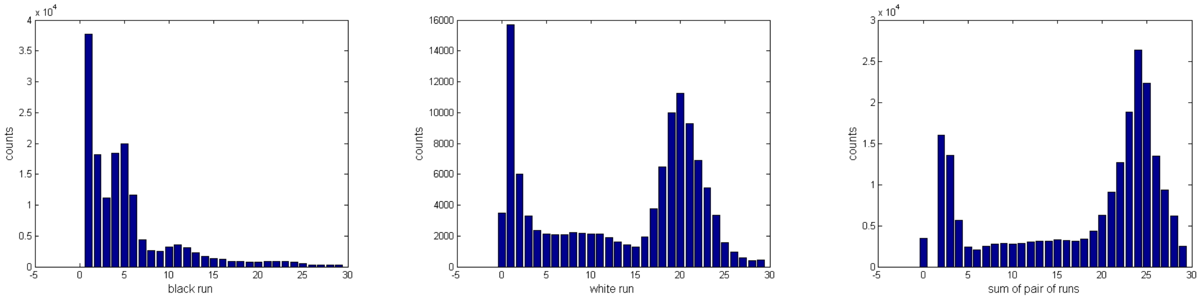
[5] I. Fujinaga, "Staff detection and removal," in *Visual Perception of Music Notation: On-Line and Off-Line Recognition*, S. George, Ed. Idea Group Inc., 2004, pp. 1–39.

[6] H. Kato and S. Inokuchi, *A recognition system for printed piano music*, H. S. Baird, K. Yamamoto, and H. Bunke, Eds. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 1992.

[1] http://gamera.sourceforge.net

Table I
MEAN AND MAXIMUM VALUE OF ERRORS (IN PIXELS) IN THE REFERENCE LENGTHS.

Length	Error	conventional estimation in binary images	proposed estimation in binary images	proposed estimation in gray-level images
staffline	mean	1.6	1.3	0.9
	max	4	3	2
staffspace	mean	2.7	1.3	1.0
	max	21	3	2
staffline+ staffspace	mean	2.4	0.4	0.4
	max	24	2	2

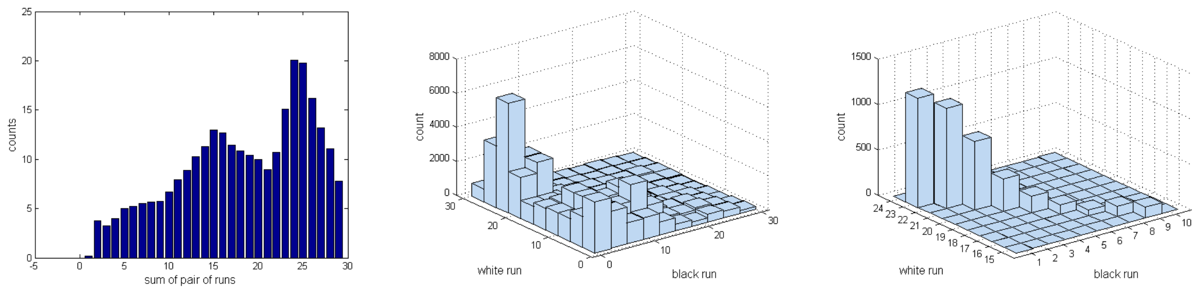


(a) Histogram of black runs.

(b) Histogram of white runs.

(c) Histogram of the sum of two consecutive runs.

Figure 3. Histogram for the music score of Figure 2 (score #17).

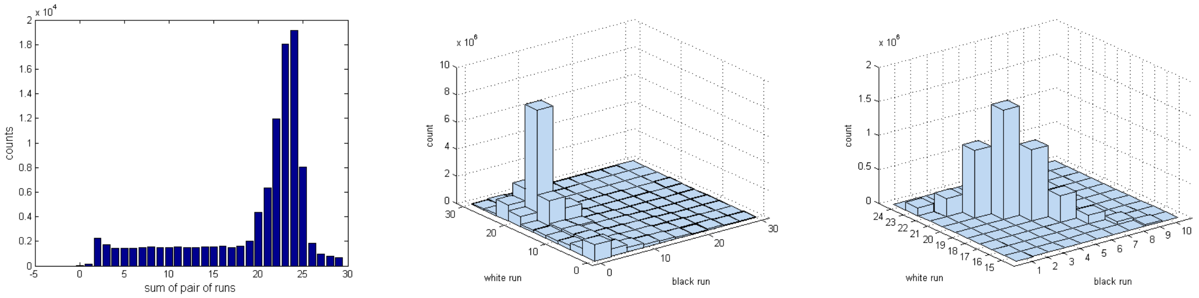


(a) Histogram of the sum of two consecutive runs. The most frequent value is 24.

(b) Histogram of the pairs (b_run, w_run).

(c) Histogram of the pairs summing 24.

Figure 5. Histograms for binarized music score #01.



(a) Histogram of the sum of two consecutive runs. The most frequent value is 24.

(b) Histogram of the pairs (b_run, w_run).

(c) Histogram of the pairs summing 24.

Figure 6. Histograms for gray-level music score #01.