Music score binarization based on domain knowledge

Telmo Pinto[1], Ana Rebelo[1], Gilson Giraldi[2], Jaime S. Cardoso[1]

[1]INESC Porto, Faculdade de Engenharia, Universidade do Porto, Portugal
[2]National Laboratory for Scientific Computing, Petrópolis, Brazil
telmotpinto@gmail.com, arebelo@inescporto.pt, gilson@lncc.br, jaime.cardoso@inescporto.pt

Abstract. Image binarization is a common operation in the pre-processing stage in most Optical Music Recognition (OMR) systems. The choice of an appropriate binarization method for handwritten music scores is a difficult problem. Several works have already evaluated the performance of existing binarization processes in diverse applications. However, no goal-directed studies for music sheets documents were carried out. This paper presents a novel binarization method based in the content knowledge of the image. The method only needs the estimation of the staffline thickness and the vertical distance between two stafflines. This information is extracted directly from the gray level music score. The proposed binarization procedure is experimentally compared with several state of the art methods.

Key words: Computer Vision, Image Processing, Optical Music Recognition

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.

After the image preprocessing (application of several techniques, e.g. binarization, noise removal, among others, to make the recognition process more robust and efficient), an OMR system can be broadly divided in three principal modules: recognition of musical symbols from a music sheet; reconstruction of the musical information to build a logical description of musical notation; construction of a musical notation model for its representation as a symbolic description of the musical sheet.

The binarization of the music score seldom justifies significant attention, with researchers invariably using some standard binarization procedure, such as the Otsu's method (e.g. [1]). Nonetheless, the development of binarization methods

specific to music scores has the potential of showing better performance than the generic counterparts and of leveraging the performance of subsequent operations.

Effective binarization of images should not only use the raw pixel information, but consider the image content as well. Unfortunately, since the binarization procedure is usually the first step of the processing system, there is usually no information available about the image content to assist the binarization procedure. A possible workaround is when we are able to extract content related information from the gray-scale image to guide the binarization procedure.

The recent work [2] on the estimation of the staffline thickness and distance without binarizing the music score, working directly in the gray-scale image, opens the door to such content aware image binarization applied to music scores, which we explore in this work.

1.1 Related Work

Different methods for image binarization have been developed and proposed in the literature. The categorization of existing techniques adopted in this work follows the survey presented in [3].

Global thresholding methods apply one threshold to the entire image. Among these techniques, the Otsu threshold selection is ranked as the best and the fastest global thresholding method [4]. It considers that the image contains two classes of pixels – foreground and background. The algorithm computes the parameter of intensity level T by maximizing the variance between the classes. This procedure considers that any point that presents an intensity equal or greater than T belongs to one class, and all the others are considered part of the other class [5]. Entropy-based methods are also very common in the image segmentation area. In [6], an image thresholding based in Tsallis entropy is proposed. The authors claim that by using Tsallis entropy they can avoid the presence of nonadditive information in some type of images that influence the segmentation operation. Edge information has also been used by several image binarization methods. In [7] the Canny edge detector was adopted: if the boundaries of actual objects in the edge image are considerably complete and closed, the binarized image can be obtained with seed filling inside the boundaries of the objects. Other works encompass similarity measures between the gray-level and the binarized images such as the minimization of fuzziness shapes [8] or the smoothed histogram via Gaussians to detect peaks and valleys [9].

In adaptive binarization methods a threshold is assigned to each pixel using local information from the image. The Bernsen's local thresholding method [10] computes the local minimum and maximum for a neighborhood around each pixel level, and uses the mean of the two as the threshold for the pixel in consideration. Niblack [11] suggested using as local information for the threshold decision the mean and the standard deviation of the pixel's neighbourhood. The works of [12, 13] applied this technique to their OMR procedures.

Although there is goal directed evaluation of binarization methods, there is no goal-directed design of binarization methods specific for certain class of images. Existing methods are generic in the sense that are agnostic of the content of the image.

2 Robust estimation of staffline 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 and the most common white-run is likely to represent the staffline spacing. Even in music scores with different staff sizes, there will be prominent peaks at the most frequent thickness and spacing. These estimates are also immune to severe rotation of the image.

In [2] the authors suggest to estimate directly the *sum* of the staffline thickness and spacing, hereafter termed `line_thickness+spacing`, since this can be robustly estimated by finding the most common sum of two consecutive vertical runs (either black run followed by white run or the reverse).

Moreover, instead of computing the most frequent peak in the histogram of the runs for a binarized image (binarized with a state-of-the-art binarization method), the authors propose 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. We propose now to use this information to guide the binarization procedure.

3 Content aware music score binarization

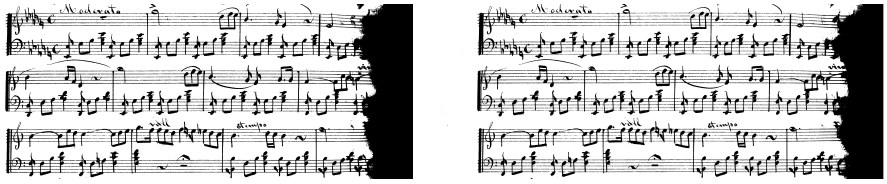
As stated in the introduction, an OMR system typically encompasses, in one of its first steps, the detection of the stafflines to facilitate the subsequent operations. A binarization method designed to maximize the number of the pairs of consecutive runs summing `line_thickness+spacing` (the peak computed over the gray-level image) will likely maximize the quality of the binarized lines. However, the direct maximization of the count of pairs of consecutive runs summing `line_thickness+spacing` could lead to a threshold value producing many, 'noisy', runs, and as a side effect, many runs at `line_thickness+spacing`. 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` but that, by chance, is the highest relative count.

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 with this condition (note that even if the integration over all thresholds does have a mode at `line_thickness+spacing`, it is possible that no individual threshold produces a histogram with mode at `line_thickness+spacing`), 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.

3.1 Using other reference lengths to guide the binarization

The same rationale used to motivate the estimation of the sum of pair of consecutive lengths, can be used to work with sets of three or more consecutive runs. However, two problems arise when proceeding that way: there is the underlying assumption that each staff have enough lines to give meaning to the consecutive runs and one starts getting less and less values to accumulate in the histogram, potentially leading to less accurate estimations.

A potentially interesting balance is estimating the sum of two times the line thickness plus the spacing, `line_2thickness+spacing`, by working with the frequencies of triplets (black run, white run, black run). This only assumes that each staff has at least two lines, but does impact the number of accumulated values, roughly halving it. The proposed content aware binarization method does not suffer any adaptation, besides the change of the reference length, `line_thickness+spacing` by `line_2thickness+spacing`. Further on in this paper we will compare the two options. In Fig. 1 we illustrate the results obtained with the proposed approach, using the two aforementioned reference lengths. One can observe that the resulting stafflines have good quality, with minor differences between the two results. Nevertheless, the original music score in this



(a) Reference length: `line_thickness+spacing`. (b) Reference length: `line_2thickness+spacing`.

Fig. 1. Result of binarizing an example of a music score.

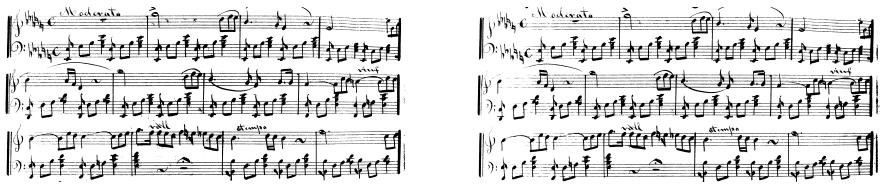
particular example is not correctly binarized with a global threshold. The digitalization of bound documents, such as books, either performed by flatbed scanners or digital cameras often yields images that exhibit a gradient-like distortion in the average colour in the region close to the book spine. In these cases, adaptive methods can show better performance.

3.2 Adaptive content aware music score binarization

Despite having been presented as a global thresholding method and having been applied it to the whole image, nothing prevents the application of the just developed ideas to a sampling window around a pixel p, effectively converting the proposed method to a local method.

As with other adaptive methods, the size of the sampling window is a key parameter. With our approach, the sampling window should be big enough to accumulate enough information (runs) to provide a proper solution. Since the typical distortions in this kind of documents are vertically oriented, the local

threshold should be constant along a column of the image. Therefore 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 calculated by interpolating the values on a set of sampled columns. In Fig. 2 we illustrate the results obtained with the proposed approach, using a window width and step size of 2% of the width of the image, and cubic polynomial interpolation. In this example, the



(a) Reference length: `line_thickness+spacing`. (b) Reference length: `line_2thickness+spacing`.

Fig. 2. Result of binarizing an example of a music score with the adaptive method.

adaptive method using the `line_thickness+spacing` reference length provided the best results, with a better staffline definition.

4 Experimental evaluation

In order to support the comparison between different binarization procedures, quantitative evaluation methods were run on a dataset of music scores. This dataset is composed of 65 handwritten scores, from 6 different authors. All the scores in the dataset were reduced to gray level information. The methods chosen try to encompass different categories of thresholding operations. Some of the algorithms tested required the input of different parameters that were obtained by experimental testing. For Sahoo's Correlation method: $Q_1 = 0.4$, $Q_2 = 1$, $Q_3 = 3$; for Tsallis entropy method: $\alpha = 2$.

For global thresholding processes, three different approaches were taken for evaluation: Difference from Reference Threshold (DRT); Misclassification Error (ME); comparison between results of staff finder algorithms applied to each binarized image. For the first method (DRT), five people were asked to choose the best possible threshold for each image. The average value of these chosen thresholds was compared to the resulting threshold value of each binarization. Ground truth versions of some scores from the dataset were also used as a comparison procedure. The Misclassification Error was defined as the difference rate between these ground truth images and the resulting images from each binarization as:

$$ME = 1 - \frac{\#(B_{bin} \cap B_{gt}) + \#(F_{bin} \cap F_{gt})}{\#B_{bin} + \#F_{bin}} \quad (1)$$

In Eq. (1) B_{bin} and F_{bin} represent the background and foreground pixels of the binarization being tested, and B_{gt} and F_{gt} the background and foreground

pixels in the reference ground truth image, respectivelly. # is the cardinality, or more precisely, the number of elements in a specific set. Since the manual binarization of the images is very time consuming, this evaluation method was applied only to ten scores chosen randomly from the complete dataset. The third technique is based on the results of staff finding algorithms applied to the binarized scores. Comparing these results, one can detect the method that will most likely produce the best outputs in the next image processing steps. The staff finding algorithms applied were Stable Path [1] and Dalitz [14]. Table 1 summarizes all the results. Both versions of the Binarization based in LIne

	Huang [8]	Khashman [15]	Kapur[16]	Sahoo [17]	Tsai [9]	Tsallis [6]	Otsu [5]	BLIST pairs	BLIST triplets
DRT: avg	48	33	50	50	29	50	19	19	29
ME: avg %	6.2	3.8	4.9	7.6	4.7	5.7	4.6	4.8	5.1
SP False: avg(std) %	2.6(5.5)	2.1(4.0)	1.4(3.4)	3.5(10.2)	2.1(4.1)	3.3(7.4)	2.0(3.4)	1.3(2.7)	1.7(3.7)
SP Missed: avg(std) %	18.0(34.5)	30.2(42.3)	27.1(42.3)	25.7(40.1)	17.0(30.3)	21.0(36.4)	8.6(20.5)	1.5(2.8)	2.8(6.3)
Dal False: avg(std) %	21.6(41.1)	3.2(7.8)	1.8(4.2)	5.4(25.6)	4.4(8.1)	2.4(6.0)	3.6(5.4)	3.2(5.0)	3.8(6.5)
Dal Missed: avg(std) %	39.6(36.9)	32.7(41.4)	31.2(42.0)	35.4(42.5)	25.4(35.0)	31.5(41.8)	18.8(31.0)	14.8(28.6)	14.9(27.4)

Table 1. Test results for various global thresholding methods, using different evaluations: difference from reference thresholds values, misclassification error (in percentage), staff detection error rates for missed and false staffs (in percentage) with Stable Path and Dalitz.

Spacing and Thickness (BLIST) method, proposed in this article, performed above average. Even so, the version that uses the pairs of runs instead of the triplets did better in the tests. This version will be considered on all the following comparisons. 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 and Misclassification Error. There are, however, two binarization techniques that get consistently better results than the others: Otsu's Method and BLIST method. The only major difference is the higher missed staff detection rate for the Otsu's algorithm.

Global methods can generally produce good outputs. Even so, for some of the scores, like those with heterogenous light distribution resulting from the digitalization process, there is no perfect threshold. In these scores, it is not possible to find a single threshold value that produces both the presence of perfectly connected staves and no occlusion of data with noise. Although staves can be correctly found in global thresholding procedures, adaptive methods can produce results with little or no loss of information.

The adaptive version of the BLIST method was implemented as described previously. The window width 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. Most adaptive algorithms tested required the input of some parameters, determined experimentally. For Bernsen: window size= 10x10 px, minimum difference in contrast = 20; for Niblack: window size = 200x200 px, k = -1; for Otsu Adaptive: window width= 2% image width; for Adaptive BLIST: window width = 2% of image width.

For the adaptive binarizations, the Misclassification Errors were all very similar. A further analysis was conducted, still based on ground truths of ten scores.

Two new error rates are presented: the Missed Object Pixel rate and the False Object Pixel, dealing with loss in object pixels and excess noise, respectively.

$$MOPx = \frac{\#F_{gt} - \#(F_{bin} \cap F_{gt})}{\#F_{gt}}, \; FOPx = \frac{\#F_{bin} - \#(F_{bin} \cap F_{gt})}{\#F_{bin}}$$

	Bernsen [10]	**Chen [7]**	**Ad BLIST**	**Niblack [11]**	**Ad Otsu**	**YB [18]**
ME: avg %	4.3	3.2	4.2	4.3	4.2	3.5
MOPx: avg %	24.6	22.5	15.6	22.5	21.7	12.4
FOPx: avg %	13.2	4.3	18.5	13.8	16.5	14.7
SP False: avg(std) %	1.3(3.0)	9.9(9.7)	2.1(5.6)	3.2(4.6)	2.7(5.9)	4.2(7.6)
SP Missed: avg(std) %	1.9(4.4)	33.0(32.8)	2.3(5.5)	14.2(23.2)	10.7(23.6)	7.9(13.8)
Dal False: avg(std) %	3.9(12.9)	3.4(5.6)	3.8(6.2)	3.1(5.1)	3.2(4.9)	3.5(6.1)
Dal Missed: avg(std) %	9.0(16.2)	17.3(27.0)	8.4(14.6)	10.2(14.1)	10.7(18.9)	7.7(10.1)

Table 2. Test results for various local thresholding methods, using different evaluations (in percentage): misclassification error, Missed Object Pixels, False Object Pixels, staff detection error rates for missed and false staffs with Stable Path and Dalitz.

Ad BLIST and YB 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. This higher noise also translates into a slightly higher false staves rate with Dalitz method. Bernsen's binarizations, although presenting the highest missed pixel rate, seem to perform well in the staff finding steps, having both the lowest missed and false staves rates.

5 Conclusion

Many binarization techniques have been proposed for digital images in the past. These methods can be applied to music scores with different rates of success although none is based on the knowledge of the content of a music score. The main contribution of this work is the introduction of a content aware binarization method for music scores. The method, based on the knowledge of the staff line thickness and spacing, extracted directly from the gray-level image, tries to find the threshold that maximizes the information content of the image, as measured by these values. We then introduced an adaptive version of our method. The basic idea of using knowledge from the image to improve the binarization operation, may apply in other areas of document image analysis, or in general image analysis.

Acknowledgments This work was partially supported by Fundação para a Ciência e a Tecnologia (FCT) - Portugal through projects PTDC/EIA/71225/2006 and SFRH/BD/60359/2009. The authors thank Prof. Paulo S. S. Rodrigues for providing the Matlab implementation of the method based in Tsallis entropy.

References

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

2. Cardoso, J.S., Rebelo, A.: Robust staffline thickness and distance estimation in binary and gray-level music scores. International Conference on Pattern Recognition **0** (2010) 1856–1859
3. Sezgin, M., Sankur, B.: Survey over image thresholding techniques and quantitative performance evaluation. Journal of Electronic Imaging **13**(1) (2004) 146–165
4. Trier, O.D., Taxt, T.: Evaluation of binarization methods for document images. IEEE Transactions on Pattern Analysis and Machine Intelligence **17**(3) (Mar 1995) 312–315
5. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man and Cybernetics **9**(1) (1979) 62–66
6. de Albuquerque, M.P., Esquef, I.A., Mello, A.R.G.: Image thresholding using tsallis entropy. Pattern Recognition Letters **25**(9) (2004) 1059–1065
7. Chen, Q., sen Sun, Q., Heng, P.A., shen Xia, D.: A double-threshold image binarization method based on edge detector. Pattern Recognition **41**(4) (2008) 1254–1267
8. Huang, L.K., Wang, M.J.J.: Image thresholding by minimizing the measures of fuzziness. Pattern Recognition **28**(1) (January 1995) 41–51
9. Tsai, D.M.: A fast thresholding selection procedure for multimodal and unimodal histograms. Pattern Recognition Leters. **16**(6) (1995) 653–666
10. Bernsen, J.: Dynamic thresholding of grey-level images (1986) In W. Bieniecki and Sz, Grabowski, Multi-pass approach to adaptive thresholding based image segmentation. Proceedings of the 8th. International IEEE Conference CADSM (2005).
11. Niblack, W.: An introduction to digital image processing (1986) In Graham Leedham and Chen Yan and Kalyan Takru and Joie Hadi Nata Tan and Li Mian, Comparison of Some Thresholding Algorithms for Text/Background Segmentation in Difficult Document Images. Proceedings of the Seventh International Conference on Document Analysis and Recognition (2003).
12. Fornés, A., Lladós, J., Sánchez, G., Bunke, H.: Writer identification in old handwritten music scores. In: DAS '08: Proceedings of the 2008 The Eighth IAPR International Workshop on Document Analysis Systems, Washington, DC, USA, IEEE Computer Society (2008) 347–353.
13. Fornés, A., Lladós, J., Sánchez, G., Bunke, H.: On the use of textural features for writer identification in old handwritten music scores. In: ICDAR '09: Proceedings of the 2009 10th International Conference on Document Analysis and Recognition, Washington, DC, USA, IEEE Computer Society (2009) 996–1000.
14. Dalitz, C., Droettboom, M., Czerwinski, B., Fujigana, I.: A comparative study of staff removal algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence **30** (2008) 753–766
15. Khashman, A., Sekeroglu, B.: A novel thresholding method for text separation and document enhancement. Proceedings of the 11th Panhellenic Conference on Informatics (PCI 2007) (May 2007)
16. Kapur, J.N., Sahoo, P.K., Wong, A.K.C.: A new method for gray-level picture thresholding using the entropy of the histogram. Computer Vision, Graphics, and Image Processing **29**(3) (March 1985) 273–285
17. Sahoo, P.K., Wilkins, C., Yeager, J.: Threshold selection using renyi's entropy. Pattern Recognition **30**(1) (1997) 71–84
18. Yanowitz, S., Bruckstein, A.: A new method for image segmentation. In: Computer Vision, Graphics, and Image Processing. Volume 46. (Apr 1989) 82–95