Staff line Detection and Removal in the Grayscale Domain

Ana Rebelo and Jaime S. Cardoso, Senior Member, IEEE

INESC TEC (formerly INESC Porto) and Faculdade de Engenharia, Universidade do Porto, Email: {arebelo,jaime.cardoso}@inescporto.pt

Abstract—The detection of staff lines is the first step of most Optical Music Recognition (OMR) systems. Its great significance derives from the ease with which we can then proceed with the extraction of musical symbols. All OMR tasks are usually achieved using binary images by setting thresholds that can be local or global. These techniques however, may remove relevant information of the music sheet and introduce artifacts which will degrade results in the later stages of the process. It arises therefore a need to create a method that reduces the loss of information due to the binarization. The baseline for the methodology proposed in this paper follows the shortest path algorithm proposed in [1]. The concept of strong staff pixels (SSP's), which is a set of pixels with a high probability of belonging to a staff line, is proposed to guide the cost function. The SSP allows to overcome the results of the binary based detection and to generalize the binary framework to grayscale music scores. The proposed methodology achieves good results.

I. INTRODUCTION

Over the last years, several researchers have been trying to overcome the lack of symbolically-represented music. Such representations of music scores enable operations such as search, retrieval and analysis. The unique way to do this is through an Optical Music Recognition (OMR) system. The increase of music libraries, some of them providing to everyone free music scores, such as the International Music Score Library Project (IMSLP)[1], promotes and encourages new challenges to the development in the OMR field.

After some image preprocessing (which may include several techniques, e.g. binarization, noise removal, blurring, de-skewing, amongst others, to make the recognition process more robust and efficient), an OMR system typically starts with a module for staff lines detection and removal to obtain an image containing only the musical symbols, where the recognition of the symbols is facilitated.

Currently, all the OMR methods need to convert the image from grayscale to binary right in the beginning of the process, facilitating the following tasks by reducing the amount of data they need to process. As a result of this, a possible loss of valuable information can happen, making unclear the advantages of binarization in the complete OMR process. Low paper quality or gradient effect on the illumination, very common in handwritten music scores, causes unsatisfactory binarization results.

The line detection and removal methods in this paper are tailored for grayscale images of handwritten music scores. The proposed paradigm uses the image as a graph, where the cost corresponding to each pixel captures the difference in luminance levels between staff lines, symbols and background present in the images.

II. STATE OF THE ART

Most of the proposed algorithms in OMR for line detection, symbol segmentation and recognition have been used for binary images rather than for grayscale or color images (e.g. [2], [3], [1], [4], [5]). This is mainly caused by the simplicity of the process and also by the typical binary nature of music scores. To the best of our knowledge, only two works have been presented to the scientific community to process grayscale images of music scores [6], [7]. In [6] an estimation of the sum of the staff line thickness (*staffline_height*) and the vertical line distance within the same staff (*staffspace_height*) was suggested, by finding the most common sum of two consecutive vertical runs (either black run followed by white run or the reverse). This work extends and improves the standard practice [2] of estimating the same reference values on binary images: the most frequent black estimates *staffline_height* and the most frequent white estimates *staffspace_height*. In [7] the Swarm Intelligence algorithm was used to detect the staff lines patterns. However, this procedure was only tested for a fragment of the original image due to the high computational complexity.

Most of the staffs are a set of five horizontal lines and four spaces which determine the musical pitch of a note on the sheet music. These lines are the longest symbols of the music score and are essential for the interpretation of music. Depending on the intended effect, the music symbols are placed vertically on the staff according to their corresponding pitch or function. Nevertheless, the staff lines, connecting and overlapping almost all musical objects, impose serious problems in an OMR system. Therefore, they are usually detected and removed in order to simplify the following OMR tasks.

One of the first strategies for staff lines detection uses the horizontal projection of the black pixels of the image [8]. Assuming that lines are straight and horizontal, the approach finds local maxima to represent the line positions. To deal with non horizontal lines, the approach computes several horizontal projections with different image rotation angles, keeping the image where the local maximum is higher. Other works combine different projections to improve the basic approach [9], [10]. A technique based on a set of *staff segments* was proposed by [11], [12], with procedures for linking two segments horizontally and vertically and merging two overlapped segments. Dalitz et.al [13] is a generalization and an improvement of this method. A rule-based classification of thin horizontal line segments [14], and line tracing [5] are also other processes for staff lines detection.

---

[1]http://imslp.org/

CPS
Conference Publishing Services

The usage of dynamic programming and the concept of graphs allowed the application of the shortest path method to the detection of the staff lines. In [1] this concept was developed for binary images, stating that if we seek the shortest path between the two opposite vertical margins of a music score and consider that the cost of a path through black pixels is lower than a path through white pixels, the shortest paths will be along staff lines.

In this framework, the design of the cost function is pivotal. In here, we extend the framework to grayscale music scores building on two novel insights. First, we introduce the concept of strong staff pixel (SSP), which by itself allows already improving the results of the binary based detection; second, making use of SSPs, we generalize the binary framework to grayscale music scores.

III. BACKGROUND

As already mentioned, the baseline for the methodology proposed in this paper follows the shortest path algorithm proposed in [1]. The goal is to have a weight function constructed to favor paths through staff-pixels – see Fig. 1 for an example. The process starts by computing two important reference values: the staff line thickness (*staffline_height*) and the vertical distance between two staff lines (*staffspace_height*).
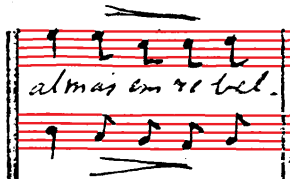

Fig. 1: An example of a music score with staff-pixels represented by the red color.

In [6] the authors proposed a method to estimate these values, both for binary and grayscale images, using the run-length encoding (RLE) technique. When each column of a binary image is encoded with RLE, the most frequent sum of two consecutive vertical runs (either black run followed by white run or the reverse) is selected as the estimation of *staffspace_height*+*staffline_height*. Then, using this information, the 2D histogram is computed in order to obtain the individual values *staffspace_height* and *staffline_height*. This method is extended to grayscale images by accumulating the runs' frequency over all the binary images that would be obtained by varying the global threshold from a low to a high limit.

After the detection of the reference values, the edges' weights are estimated [1] (as discussed later). Afterwards, and during the main cycle, the algorithm successively finds the shortest paths between the left and right margins, using the edges' weights previously estimated. To stop the iterative staff line search, two validation rules were applied: a path is discarded if it does not have a percentage of black pixels above the median percentage of blackness of all lines found in the first iteration of the main cycle; and a path is discarded if its shape differs too much from the shape of the line with median blackness. If the path detected does not pass the checking, the iterative search is stopped. After the main search step, valid staff lines are post-processed. The algorithm eliminates

spurious lines and cluster them in staves, and at the end lines are smoothed and can be trimmed.

The shortest path algorithm [1] depends on the design of the edges' weights function. The main idea was to assign a low cost to the edge between two black pixels and high cost otherwise, and, at the same time, to distinguish between black pixels in the staff lines from black pixels in the music symbols, penalizing the latter and favoring the former. In this manner, the authors incorporated several factors in the cost function according to some prior knowledge about a music score. A term benefiting edges of incident black pixels of short vertical run of black pixels is included. In this case, if a black pixel is part of a short vertical run of black pixels, then it is more likely to be part of a staff line rather than of a symbol. A penalizing term is incorporated in the edges where the nearest vertical run of black pixels on the same column is excessively far from the vertical run of black pixels containing the current black pixel. In this case, the authors are considering that a staff line is likely to have another staff line at roughly *staffspace_height* pixels (assuming that staves have at least two lines), otherwise the pixels are more likely to belong to a symbol (probably a ligature) rather than to staff line.

IV. STRONG STAFF PIXELS

We start by introducing the concept of strong staff-pixel in binary images. The integration of this concept in the weight function has already the potential to improve the performance of the detection methods. Further processing can be performed based on the knowledge of the position of these pixels and the search for the staff lines could be guided through them. Afterwards, we extend the strong staff-pixel definition for grayscale images.

A. Strong Staff-Pixels

Intuitively, a strong staff-pixel is a pixel with a high probability of belonging to a staff line. The decision of accepting/rejecting a pixel as a strong staff-pixel is not made independently for each pixel but builds on the run length encoding previously mentioned. All pixels in the same run will either be considered as staff pixels or not staff pixels. The decision approach consists in scanning the run-length encoding of each column of the binary image of the music sheet in order to find black runs of *staffline_height* pixels followed or preceded by a white run of *staffspace_height* pixels. See Fig. 2 for visualization on an example. The pixels in the black runs that meet this condition form the set of Strong Staff-pixels (SSP) – see Fig. 3.

In grayscale domain, instead of computing the previously explained method for just only one binary image obtained by a state-of-the-art binarization technique, the scanning RLE is performed for every possible binary image by varying the threshold from a low to a high limit. In this manner, the final set of SSP will be formed by accumulating sets of black runs with an height equal to *staffline_height* that are followed or preceded by a white run of *staffspace_height* pixels.

B. Improved Weight Function in the Binary Domain

The pseudo-code for the weight function is provided in Listing 1. The first version of this function was already presented in Section III – for more details please see [1]. The main
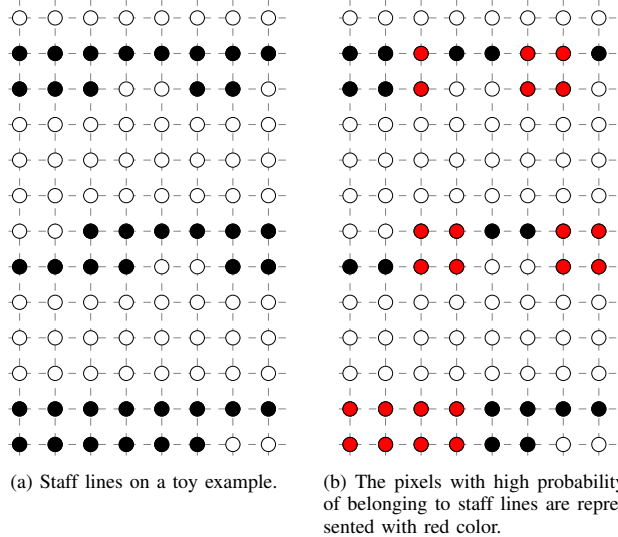
(a) Staff lines on a toy example.

(b) The pixels with high probability of belonging to staff lines are represented with red color.

Fig. 2: An illustrative example of the strong staff-pixel algorithm. The *staffspace_height* is 3 and *staffline_height* is 2.

(a) Original music score from the Fig. 1.

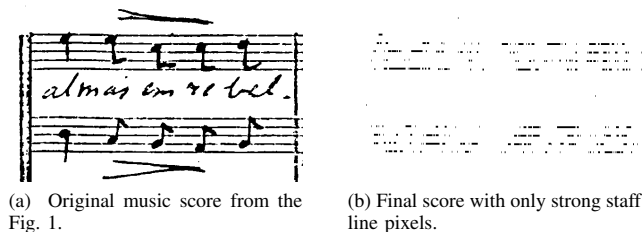(b) Final score with only strong staff line pixels.

Fig. 3: Estimation of strong staff-pixels using the proposed methodology.

difference is in the use of the concept of SSP to penalize or to favour the current pixel. Hence, if the current pixel belongs to the set of SSP ($SSP = 1$) then a benefit term is incorporated to the weight. Please note that because of the way SSP is built this current pixel is always a black pixel and its vertical run of black pixels has always an height equal to *staffspace_height*. Otherwise, the black pixels ($pixelValue = 0$) that do not belong to this group ($SSP = 0$) and belong to a vertical run of black pixels higher than the *staffspace_height* are penalized. In here, with the information given by the SSP, it is possible to distinguish with more certainty the black pixels of staff lines from the ones of music symbols. The last restriction is equal to the first version of the weight function and is related to the edges where the nearest vertical run of black pixels on the same column is excessively far from the vertical run of black pixels containing the current black pixel.

Remembering the reader, in the first version, a term benefiting edges of incident black pixels of short vertical run of black pixels was included. In noisy images this simple restriction could lead to benefit broken music symbols or noise. With this new cost function only the SSP are considered. Nevertheless, since we do not have certainty about the black pixels in that condition they were not penalized.

```
WeightFunction(SSP, pixelValue, vRun1,
 vRun2, nearestVRun1, nearestVRun2, NeighbourhoodType)
{
  weight = baseWeight(pixelValue, NeighbourhoodType);
  if(SSP==1)
   weight = weight - delta;
  if( (vRun1>STAFFLINEHEIGHT)
    OR(vRun2>STAFFLINEHEIGHT)
    AND (SSP==0) AND (pixelValue==0) )
    weight = weight + delta;
  if( (nearestVRun1>=STAFFSPACEHEIGHT+STAFFLINEHEIGHT)
    OR(nearestVRun2>=STAFFSPACEHEIGHT+STAFFLINEHEIGHT))
    weight = weight + delta;
  return weight;
}
```

Listing 1: Pseudo-code for the weight Function. The base weight was set to 4 on black pixels and 8 on white pixels for 4-neighbourhoods and to 6 and 12 on for 8-neighbourhoods. The *delta* penalizing term in the weight function was set to 1. For efficiency, weights were designed with integer values.

V. GRAYSCALE STAFF LINE DETECTION

For binary staffs, it was sufficient to assign a low cost for incident black pixels and a high cost for the white pixels. Now, instead of only two values ($\{0, 1\}$, background and foreground) we have a range of values in $[0, 1]$. Therefore, the weight function has to be generalized to this new domain.

Our proposed weight function is based on a sigmoid function where the parameters are chosen to favor the luminance levels of staff – see Eq. 1. The rationale is to extend the penalization of pixels in the binary procedure that can be seen as a step function (pixels binarized to white are assigned a high cost; pixels binarized to black are assigned a low cost). With this in mind, the sigmoid function selected for 4-neighbourhoods was the following:

$$w(p) = b/(1 + \exp(-(p - \alpha)/\beta)) + a \qquad (1)$$

where p is the minimum of the values of the two pixels incident with the edge. In the case of 8-neighbourhoods the weight is $\sqrt{2}$ times that value. An example of a plot for the function can be seen in Fig. 4.

The a and b are normalization factors for the values range of our cost function and they were both set to 4; α was set as the threshold value from the BLIST binarization [15].

The β parameter is inversely related to the slope of the sigmoid at the center of symmetry α. Since different scores can have different dispersion of values within the staff lines, one chose to set the β proportional to the standard deviation, *staffvalue_std*, of the gray intensity levels of the SSP's. Experimentally β was set to $0.8 \times$ *staffvalue_std*. The pseudo-code for the weight function is provided in Listing 2.

The *staffvalue_std* value can also be computed in an adaptive way with the aim to overcome the differences in luminance that often occur along the same handwritten music score – see Fig. 5. Since, the sampling window should capture enough information (sets of staff lines) to provide a robust solution, a window with height equal to the height of the image and width of 100 pixels (experimentally chosen) without overlapping was selected. Hence, different *staffvalue_avg* and *staffvalue_std* values are obtained for each set of columns.
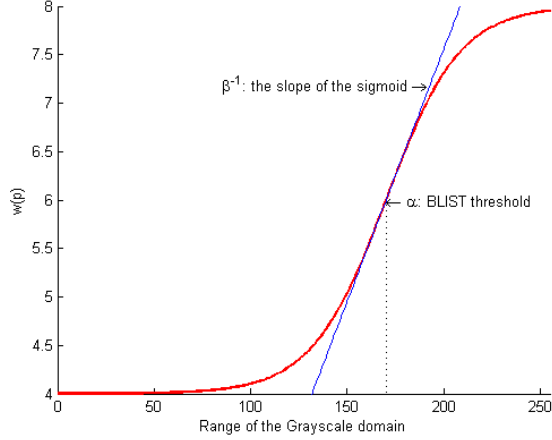
Fig. 4: The cost function $w(p)$ for the music score presented in Fig. 3a. For this example $\alpha = 170$ and $\beta = 19.2$.

Fig. 5: An example of a music score where the gray-level of staff pixels is variable from left to the right margin of the image.

VI. METRICS AND RESULTS

In this work two different evaluations were carried out: the quality of the pixels that belongs to the set of SSP and the quality of the line detection algorithm. The evaluation of the quality of the SSP was obtained comparing the resulting image with the references positions of symbols and staff lines. The evaluation of the performance of the line detection algorithm were done through the evaluation of the performance of the staff line detection and removal algorithms. A set of 76 of handwritten music scores, for which reference staff lines were manually outlined, was used.

A. SSP Evaluation

For the first experimental testing the resulting images with only staff line pixels – for instance, images as the one presented in Fig. 3 – were compared with images with only true staff lines pixels to count the number of false pixels detected. This number restricts the quality of the SSP. The number of missed pixels in this situation is not relevant provided that we have pixels in real positions of staff lines. The staff lines ground truth images are composed by the true positions of the staff lines with breaks on the symbols positions. The symbols ground truth images composed by music symbols without noise were compared to the resulting images with only staff-pixels. The aim was to check if the false positives (pixels wrongly classified as staff lines) detected by the algorithm belong to noise or to music symbols.

The results are presented on the Table I. The number of pixels wrongly classified as non-staff lines is 7% against 93%

```
WeightFunction(pixelValue1, pixelValue2, a, b, alpha, beta,
 vRun1, vRun2, nearestVRun1, nearestVRun2, NeighbourhoodType)
{
  value = min(pixelValue1, pixelValue2);
  weight = sigmoidFc(value,a,b,alpha,beta,NeighbourhoodType);
  if( (vRun1<=STAFFLINEHEIGHT)
    OR(vRun2<=STAFFLINEHEIGHT) AND value <= alpha)
    weight = weight - delta;
  if( (nearestVRun1>=STAFFSPACEHEIGHT+STAFFLINEHEIGHT)
    OR(nearestVRun2>=STAFFSPACEHEIGHT+STAFFLINEHEIGHT))
    weight = weight + delta;
  return weight;
}
```

Listing 2: Pseudo-code for the weight Function. The sigmoidFC is the Equation 1. The $delta$ penalizing term in the weight function was set to 1. For efficiency, weights were designed with integer values.

for pixels classified as staff lines, which represents a very satisfactory result. Usually, in an music score, the most of the symbols appear on the staff lines and noise could be anywhere. As the goal is to follow the staff lines as close as possible it is better to classify staff-pixels as symbols than as noise.

	Staff lines	Symbols	Noise
average	93%	4%	3%
standard deviation	4%	2%	4%

TABLE I: Pixels classification obtained by the staff-pixel algorithm.

B. Staff Line Detection Evaluation

For the evaluation of the staff line detection algorithm two metrics were considered: the number of unmatched detected staff lines (false positive) and the number of unmatched reference staff lines (missed to detect). They were computed by the average Euclidian distance between each reference staff line and each actually detected staff line, where the matching problem on the resulting bipartite graph was solved by minimizing the distance [1].

Both binary and grayscale domains were compared. The SSP approach for binary images was also compared with the first weight function proposed in [6].

The binarization of the images was carried out using the BLIST method [15]. The outcome of the algorithms can be seen on Table II. When the SSP was utilised to guide the searching of the staff lines the results were better. In the case of grayscale images the staff line detection algorithm obtained better results. On Fig. 6, it is possible to see an example of a music score where the binary method (BLIST) fails to find all the staff lines and the grayscale method does not.

|	Weight function	False detection rate	Miss detection rate
BLIST	[6]	1.0 (2,0)	1.2 (2.3)
BLIST	SSP	0.8 (2.1)	0.9 (2.2)
Grayscale	Sigmoid (global *staffvalue_std*)	0.7 (2.2)	0.9 (2.8)
Grayscale	Sigmoid (Local *staffvalue_std*)	0.7 (1.8)	0.8 (2.0)

TABLE II: Detection performance in percentage: average (standard deviation).

In a further experience, we used a staff line removal operation to test our procedures [1]. In the binary domain,

(a) Using BLIST.
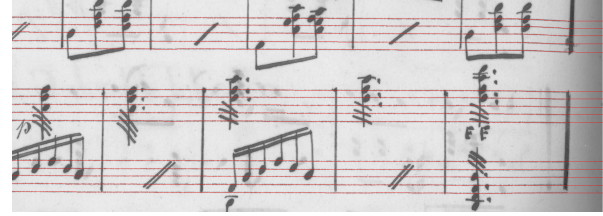
(b) Using grayscale image.

Fig. 6: Staff lines detected of a music score from our database.

for each pixel of a detected staff line the procedure removes every vertical black run thinner than $2\times$ *staffline_height*. It is important to stress that RLE coding was used to obtain the run sequences. In order to extend this method to the grayscale domain, the BLIST threshold [15] was used for each pixel of a detected staff line to compute the vertical black runs. Once again, to avoid issues related to the changes in the luminance along of a image this binary threshold value was obtained for a sets of columns, as already explained in the previous section.

The same error metric from [16] was adopted. This metric evaluates the number of misclassified staff-pixels and the number of misclassified non-staff-pixel in relation to the number of all staff-pixels. It is defined by:

$$E.R. = 100 \times \frac{\#misclassified\ sp + \#misclassified\ non\ sp}{\#all\ sp + \#all\ non\ sp} \quad (2)$$

The results, using the best process for both domains, are shown on Table III. The grayscale images obtained the best accuracy, proving the advantage in relation to binarization.

Domain	mean error	standard deviation error
Binary	0.13	0.20
Grayscale	0.06	0.08

TABLE III: Errors on staff line removal.

VII. CONCLUSION

The purpose of working on the gray-level domain was to preserve the information that where otherwise lost in the binarization process (common on OMR systems), making this method a more robust solution to work with handwritten music scores with corrupted backgrounds due to problems during the digitization, damaged paper based documents, among several others anomalies that could be problematic in the following steps of an OMR system. In this paper, the concept of SSP was introduced in order to guide the staff line detection algorithm proposed in [1]. The new methodology improves the results of the binary based detection and made possible to generalize the framework to grayscale domain. In here, a new cost function capturing the luminance difference between levels of staff lines, symbols and background was proposed.

ACKNOWLEDGMENT

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*, vol. 31, no. 6, pp. 1134–1139, 2009.

[2] 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.

[3] L. Pugin, "Optical music recognition of early typographic prints using Hidden Markov Models," in *Proceedings of the International Society for Music Information Retrieval*, 2006, pp. 53–56.

[4] A. Rebelo, G. Capela, and J. S. Cardoso, "Optical recognition of music symbols: A comparative study," *International Journal on Document Analysis and Recognition*, vol. 13, pp. 19–31, 2010.

[5] L. J. Tardón, S. Sammartino, I. Barbancho, V. Gómez, and A. Oliver, "Optical music recognition for scores written in white mensural notation," *EURASIP Journal on Image and Video Processing*, vol. 2009, pp. 6:3–6:3, February 2009.

[6] J. S. Cardoso and A. Rebelo, "Robust staffline thickness and distance estimation in binary and gray-level music scores," in *20th International Conference on Pattern Recognition*, August 2010, pp. 1856–1859.

[7] W. Piatkowska, L. Nowak, M. Pawlowski, and M. Ogorzalek, "Stafflines pattern detection using the swarm intelligence algorithm," in *Computer Vision and Graphics*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2012, vol. 7594, pp. 557–564.

[8] 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.

[9] F. Rossant and I. Bloch, "Robust and adaptive OMR system including fuzzy modeling, fusion of musical rules, and possible error detection," *EURASIP Journal on Applied Signal Processing*, vol. 2007, no. 1, pp. 160–160, 2007.

[10] P. Bellini, I. Bruno, and P. Nesi, "Optical music recognition: Architecture and algorithms," in *Interactive Multimedia Music Technologies*. Hershey: IGI Global, 2008, pp. 80–110.

[11] 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.

[12] M. Szwoch, "A robust detector for distorted music staves," in *Computer Analysis of Images and Patterns*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2005, pp. 701–708.

[13] 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.

[14] J. V. Mahoney, "Automatic analysis of music score images," 1982, in D. Blostein and H. Baird. A Critical Survey of Music Image Analysis. In *Structured Document Image Analysis*, pages 405–434, Springer-Verlag, Heidelberg, 1992.

[15] T. Pinto, A. Rebelo, G. Giraldi, and J. S. Cardoso, "Music score binarization based on domain knowledge", booktitle = Proceedings of Iberian Conference on Pattern Recognition and Image Analysis, pages = 700–708, year = 2011,."

[16] A. Fornés, A. Dutta, A. Gordo, and J. Lladós, "CVC-MUSCIMA: a ground truth of handwritten music score images for writer identification and staff removal," *International Journal on Document Analysis and Recognition*, vol. 15, pp. 243–251, 2012.