The ICDAR 2011 Music Scores Competition: Staff Removal and Writer Identification

Alicia Fornés, Anjan Dutta, Albert Gordo and Josep Lladós
Computer Vision Center - Dept. of Computer Science
Universitat Autònoma de Barcelona, Edifici O, 08193
Bellaterra, Spain
Email: {afornes,adutta,agordo,josep}@cvc.uab.es

***Abstract*—In the last years, there has been a growing interest in the analysis of handwritten music scores. In this sense, our goal has been to foster the interest in the analysis of handwritten music scores by the proposal of two different competitions: Staff removal and Writer Identification. Both competitions have been tested on the CVC-MUSCIMA database: a ground-truth of handwritten music score images. This paper describes the competition details, including the dataset and ground-truth, the evaluation metrics, and a short description of the participants, their methods, and the obtained results.**

***Keywords*-competition; music scores; writer identification; staff removal;**

I. Introduction

In the last years, there has been a growing interest in the analysis of handwritten music scores [1], [2], [3]. In this context, the focus of interest is two-fold: the recognition of handwritten music scores (Optical Music Recognition), and the identification (or verification) of the authorship of an anonymous music score.

In the Optical Music Recognition systems, staff removal algorithms have attracted many researchers [5], [7], since a good detection and removal of the staff lines will allow the correct isolation and segmentation of the musical symbols, and consequently, will ease the correct recognition and classification of the music symbols.

Nowadays, musicologists must work very hard to identify the writer of an unknown manuscript. In fact, they do not only perform a musicological analysis of the composition (melody, harmony, rhythm, etc), but also analyze the handwriting style of the manuscript. In this sense, writer identification can be performed by analyzing the shape of the hand-drawn music symbols (e.g. music notes, clefs, accidentals, rests, etc), because it has been shown that the author's handwriting style that characterizes a piece of text is also present in a graphic document.

In order to foster the interest in the analysis of handwritten music scores, we have proposed at ICDAR and GREC (International Workshop on Graphics Recognition) two different competitions: Staff removal and Writer Identification. Both competitions have been tested on the CVC-MUSCIMA database [4] of handwritten music score images.

This paper describes the competition details, including the dataset, ground-truth, and the evaluation metrics. Afterwards, we shortly describe the 16 submitted methods (8 methods in each competition) and analyze the obtained results. The rest of the paper is organized as follows. Section 2 describes the dataset and the staff distortions applied. Section 3 describes the staff removal competition, whereas the writer identification competition is described in Section 4. Finally, concluding remarks are described in Section 5.

II. Database

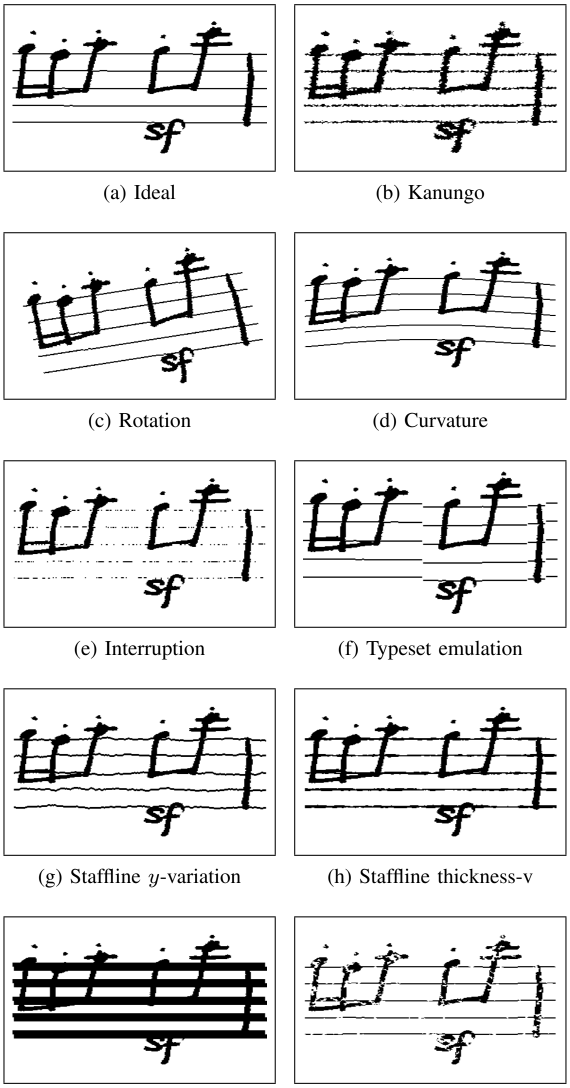
The CVC-MUSCIMA database has been designed for musical scores analysis and recognition, but in the first stage the ground truth created has been addressed to staff removal and writer identification. This database is fully described in [4] and available in the website: http://www.cvc.uab.es/cvcmuscima. It consists of 1,000 handwritten music score images, written by 50 different musicians. All the 50 writers are adult musicians in order to ensure that they have their own characteristic handwriting music style. Each writer has transcribed exactly the same 20 music pages, using the same pen and the same kind of music paper.

III. Staff Removal Competition

For testing the robustness of the staff removal algorithms, we have applied the following distortion models (see Fig.1) to the original images: degradation with Kanungo noise, rotation, curvature, staffline interruption, typeset emulation, staffline y-variation, staffline thickness ratio, staffline thickness variation and white speckles. Two of these models (staffline y-variation and staffline thickness variation) are applied twice with different parameters. See [4] for details.

As a result, we have obtained 11,000 distorted images, with together with the originals yield a total of 12,000 images. For the staff removal competition the entire dataset is equally divided into two parts, of which the first 50% of the images (500 images x 12 variations = 6000 images) will be used as training the algorithms and the other 50% (6000 images) of the images will be used for testing them.

1520-5363/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDAR.2011.300



(a) Ideal (b) Kanungo (c) Rotation (d) Curvature (e) Interruption (f) Typeset emulation (g) Staffline y-variation (h) Staffline thickness-v (i) Staffline thickness Ratio (j) White speckles

Figure 1: Examples of Staff deformations.

A. Participants

In this subsection, we will shortly describe the methods submitted by the participants.

1) ISI01: This system was submitted by Jit Ray Chowdhury and Umapada Pal from the Computer Vision and Pattern Recognition Unit of the Indian Statistical Institute, Kolkata, India. The authors submitted two versions of the algorithm:

- ISI01-Rob: First, the images are thinned and, by analyzing the thinned portions, the input images are automatically categorized in two groups: (a) images containing straight staff line and (b) other non-straight or curved staff-lines. Images containing straight staff lines are further divided into horizontal staff lines and non-horizontal straight lines. Next, staff lines are detected based on the characteristics of each group. Some smoothing techniques are also utilized to get better accuracy. The staff line detections methods developed here can be considered as passing a ring on a wire (here wire can be considered as staff-line). If there is any obstacle like music score the obstacle portions is retained or deleted based on some measures. For staff-line detection the authors computed staff line height, staff space height, vertical positional variance of the pixels of thinned lines, etc. These parameters guided their system to detect the staff line part efficiently.
- ISI01-HA: The second method corresponds to a second version of the previous method, where the parameters were set to minimize average error rate but without any restriction for maximum error rate.

2) INP02: These systems were submitted by Ana Rebelo and Jaime S. Cardoso from the Institute for Systems and Computer Engineering of Porto, Portugal. The authors propose a graph-theoretic framework where the staff line is the result of a global optimization problem, which is fully described in [5]. The authors submitted two methods:

- INP02-SP: The staff line algorithm uses the image as a graph, where the staff lines result as connected paths between the two lateral margins of the image. A staff line can be considered a connected path from the left side to the right side of the music score. The main cycle of the methodology consists in successively finding the stable paths between the left and right margins, adding the paths found to the list of staff lines, and erasing them from the image. To stop the iterative staff line search, a sequence of rules is used to validate the stable paths found; if none of them passes the checking, the iterative search is stopped. A path is discarded if it does not have a percentage of black pixels above a fixed threshold. Likewise, a path is discarded if its shape differs too much from the shape of the line with median blackness. After the main search step, valid staff lines are post-processed. The algorithm eliminates spurious lines and cluster them in staves. Finally, lines are smoothed and can be trimmed.
- INP02-SPTrim: In this version, the aim is to eliminate the initial white pixels of the paths. Hence, for each staff, a sequence of median colours is computed as follows: for each column, the median of the colours (black and white values) of the lines is added to the sequence. Next, the trimming points are found on this sequence: starting on the centre, we traverse the sequence to the left and right until a run of $whiterun = 2staffspaceheight$ white pixels is found. The pixels between the left and right runs are kept in the staff lines. The weight function was designed to favour the black pixels of the staff lines. Hence, the function assigns

high costs for white pixels and black pixels of the music symbols.

3) NUS03: This method was submitted by Bolan Su from the School of Computing of the National University of Singapore; Shijian Lu from the Institute for Infocomm Research, Singapure; Umapada Pal from the Computer Vision and Pattern Recognition Unit of the Indian Statistical Institute, Kolkata, India; and Chew-Lim Tan from the School of Computing of the National University of Singapore.

The method consists in the following: First the staff height and staff space are estimated using the histogram of vertical run length. Those staff lines are assumed parallel, then the estimated staff height and space are used to predict the lines' direction and fit an approximate staff line curve for each image. The fitted staff line curve can be used to identify the actual location of staff lines on the image. Then those pixels who belong to staff lines are removed.

4) NUG04: These systems were submitted by Christoph Dalitz and Andreas Kitzig from the Niederrhein University of Applied Sciences, Institute for Pattern Recognition (iPattern), Krefeld, Germany. The authors submitted three different systems:

- NUG04-Fuji: The method identifies long horizontal runs as staffline candidates. To allow for possible curvature, the image is in a preprocessing step deskewed by alignment of vertical strips based on their projection correlation. This however only works for a very limited range of curvature or rotation. For more details on the Fujinaga's approach, see [6]. The source code is available in the website: *http://music-staves.sourceforge.net/* (class $MusicStaves\_rl\_fujinaga$).
- NUG04-LTr: The method simply removes all vertical runs shorter than $2 * staffline - height$ around a found staff line. The $staffline - height$ is measured as the most frequent black vertical runlength. The staff finding is done by vertically thinning long horizontal runs with an average blackness above a certain threshold, vertically linking these filaments based upon their vertical distance and then identifying staff systems as connected subgraphs. The first step of identifying long horizontal dark windows makes this method inappropriate for strongly curved stafflines. For more details, see [7] (Section 3.1, method "Linetrack Height" with the staff finder described at the end of section 2). The source code is available in the website: *http://music-staves.sourceforge.net/* (class $MusicStaves\_linetracking$).
- NUG04-Skel: The method directly discriminates staff segments from musical symbols. It is based on splitting the skeleton image at branching and corner points and building a graph with vertical and horizontal links from those segments fulfilling heuristic rules that make them likely to be staffline segments. As the horizontal linking is based on extrapolation, this method fails for heavily curved stafflines. For more details, see [7] (Section 3.4). The source code is available in the website: *http://music-staves.sourceforge.net/* (class $MusicStaves\_skeleton$).

B. Metrics and Results

The performance of the algorithms was measured based on pixel based metric. Here the staff removal is considered as a two-class classification problem at the pixel level. The error rate of classification for each of the images ranges from 0 to 100, and was computed as

$$\text{E.R.} = 100 \cdot \frac{\#\text{misclassified } sp + \#\text{misclassified non } sp}{\#\text{all } sp + \#\text{all non } sp} \quad (1)$$

where # means "number of" and *sp* means "staff pixels". So lower being the error rate, better the performance.

Since it may occur that one system obtains very good results but rejects many images, the systems provided by the participants have been evaluated in two ways:

- Error rate without rejection: The error rate of the images that the system could evaluate. Thus, the rejected images are not included here.
- Error rate with rejection: The error rate is computed taking into account all the set of images. Thus, the rejected images are included with an E.R.=100%.

The results of the different methods are shown in Table I. Most methods have an error rate without rejection between 1.9 and 2.8, being ISI01-HA the one which obtains better results in most cases, and also without rejecting any image. Concerning the rejected images, one can see how the NUS03 method has lower Error Rate than the INP02 methods, but discards all the *Thick* distorted images. In this sense, it is important to note that some severe distortions (such as Interruption or Thickness) make the staff detection very difficult, and consequently, most images are rejected by the systems (in many cases, all the images are discarded).

IV. WRITER IDENTIFICATION COMPETITION

For the writer identification competition, the dataset is equally divided in two parts, where 500 images (10 images from each writer) were used for training, and 500 images were used for testing. We have provided images without the staff lines (see Fig.2), because they are particularly useful for writer identification: since most writer identification methods remove the staff lines in the preprocessing stage, this eases the publication of results which are not dependent on the performance of the particular staff removal technique applied. Moreover, these images make easy the participation of researchers that do not work on staff removal.

A. Participants

In this subsection, we will shortly describe the methods submitted by the participants.

Table I: Staff Removal results. Error Rate (E.R.) in % for each one of the 12 distortions. We show the Error Rate with and without rejection. In case of the Error rate without rejection (No Rej.), we also show the number # of rejected images.

Distortion	Error Rate	ISI01-Rob	ISI01-HA	INP02-SP	INP02-SPTrim	NUS03	NUG04-Fuji	NUG04-LTr	NUG04-Skel
01- Ideal	No Rej.(#) With Rej.	1.50 (0) 1.50	1.50 (0) 1.50	1.5 (0) 1.5	1.51 (0) 1.5	1.54 (0) 1.54	1.53(0) 1.53	2.08 (0) 2.08	2.11 (1) 2.31
02- Curvature	No Rej.(#) With Rej.	1.66 (0) 1.66	1.66 (0) 1.66	1.8 (0) 1.8	1.80 (0) 1.8	2.83 (0) 2.83	38.45 (3) 38.82	**– (500)** **100**	13.38 (148) 39.02
03- Interruption	No Rej.(#) With Rej.	0.92 (0) 0.92	0.91 (0) 0.91	5.16 (5) 6.10	5.19 (5) 6.14	1.04 (0) 1.04	**18.79 (499)** **99.84**	**– (500)** **100**	**– (500)** **100**
04- Kanungo	No Rej.(#) With Rej.	2.84 (0) 2.84	2.84 (0) 2.84	2.86 (0) 2.86	2.87(0) 2.87	2.91 (0) 2.91	2.84 (0) 2.84	4.33 (0) 4.33	7.93 (0) 7.93
05- Rotation	No Rej.(#) With Rej.	1.76 (0) 1.76	1.76 (0) 1.76	2.03 (0) 2.03	2.03 (0) 2.03	3.06 (0) 3.06	40.40 (8) 41.35	**– (500)** **100**	4.60 (48) 13.76
06- staffline thickness v1	No Rej.(#) With Rej.	2.44 (0) 2.44	2.17 (0) 2.17	2.70 (0) 2.70	2.71 (0) 2.71	3.38 (0) 3.38	2.53 (0) 2.53	3.74 (0) 3.74	4.14 (0) 4.14
07- staffline thickness v2	No Rej.(#) With Rej.	2.18 (0) 2.18	2.15 (0) 2.15	3.01 (0) 3.01	3.02 (0) 3.02	3.41 (0) 3.41	2.20 (0) 2.20	3.74 (0) 3.74	3.72 (0) 3.72
08- staffline y-variation v1	No Rej.(#) With Rej.	2.00 (0) 2.00	1.89 (0) 1.89	2.43 (0) 2.43	2.45 (0) 2.45	3.01 (0) 3.01	3.21 (0) 3.21	5.56 (2) 5.94	6.34 (0) 6.34
09- staffline y-variation v2	No Rej.(#) With Rej.	1.92 (0) 1.92	1.83 (0) 1.83	2.27 (0) 2.27	2.28 (0) 2.28	3.02 (0) 3.02	3.28 (0) 3.28	3.34 (2) 3.72	4.98 (0) 4.98
10-Thickness Ratio	No Rej.(#) With Rej.	2.86 (0) 2.86	2.86 (0) 2.86	6.89 (0) 6.89	6.89 (0) 6.89	**– (500)** **100**	**– (500)** **100**	10.78 (0) 10.78	15.96 (0) 15.96
11-TypeSet emulation	No Rej.(#) With Rej.	1.61 (0) 1.61	1.60 (0) 1.60	1.60 (0) 1.60	1.61 (0) 1.61	1.70 (0) 1.70	7.95 (0) 7.95	3.29 (8) 4.83	**18.41 (477)** **96.25**
12- WhiteSpec	No Rej.(#) With Rej.	1.48 (0) 1.48	1.48 (0) 1.48	1.73 (0) 1.73	1.74 (0) 1.74	2.04 (0) 2.04	1.92 (0) 1.92	1.76 (0) 1.76	6.69 (0) 6.69
Overall E.R.	No Rej.(#) With Rej.	1.93 (0) 1.93	**1.89** (0) 1.89	**2.83** (5) 2.91	2.84 (5) 2.92	**2.54 (500)** 10.66	10.37 (1010) 25.46	**4.29 (1512)** 28.41	6.87 (1174) 25.09

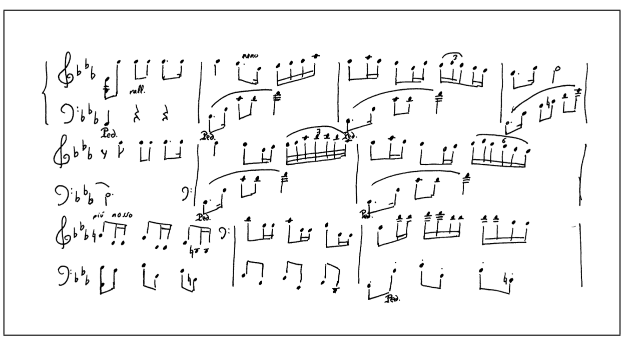


Figure 2: Example of image without staff lines.

1) PRIP02: These methods were submitted by Abdelâali Hassaïne and Somaya Al-Ma'adeed from the Pattern Recognition and Image Processing Research Group of Qatar University; and Ahmed Bouridane from Northumbria University. The authors submitted three methods:

- PRIP02-edges: The first one uses the edge-based directional probability distribution features (see [8]).
- PRIP02-grapheme: The second one uses grapheme features, and it is fully described in [9].
- PRIP02-combination: The third method combines both kinds of features, edge-based and grapheme features.

These methods have previously been applied for Arabic writer identification and for signature verification and have shown interesting results. The classification step is performed either using a logistic regression classifier or a k-nearest neighbour algorithm.

2) TUA03: These methods were submitted by Chawki Djeddi from the Mathematics and Computer Science Department of the Cheikh Larbi Tebessi University, Tebessa, Algeria; and Labiba Souici-Meslati from the LRI Laboratory, Computer Science Department of the Badji Mokhtar University, Annaba, Algeria.

The methods compute run-lengths features, which are determined on the binary image taking into consideration the pixels corresponding to the ink trace. The probability distribution of white run-lengths has been used in the writer identification experiments. There are four scanning methods: horizontal, vertical, left-diagonal and right-diagonal. The authors calculate the runs-lengths features using the grey level run-length matrices and the histogram of run-lengths is normalized and interpreted as a probability distribution. For further details, see [10].

Table II: Writer Identification results. Number of correct images and the final Writer Identification (W.I.) rate in %.

Method	Correct/Total	W.I.rate (%)
PRIP02-edges	327/500	65.4
PRIP02-grapheme	319/500	63.8
PRIP02-combination	385/500	**77.0**
TUA03-5NN	267/500	53.4
TUA03-MPL	324/500	64.8
TUA03-SVMOAA	383/500	**76.6**
TUA03-SVMOAO	333/500	66.6
TUA03-combination	352/500	70.4

For the classification step, the authors have used five different approaches:

- TUA03-5NN: A 5 nearest neighbor classifier (5-NN) with cityblock Distance Metric.
- TUA03-SVMOAO: A Support Vector Machine classifier (SVM One against one).
- TUA03-SVMOAA: A Support Vector Machine classifier (SVM One against all).
- TUA03-MLP: Multilayer perceptron (MLP).
- TUA03-Combination: A combination of the four previous classifiers: a multilayer perceptron (MLP), two Support Vector Machine classifiers (SVM One against all, SVM one against one) and a 5-NN classifier with cityblock Distance Metric. The combination rule used in their experiments is Majority Vote.

B. Metrics and Results

A musical score will be considered as correctly classified if the writer decided by the algorithm is the same as the ground-truthed one. The evaluation metric will be the Writer Identification rate $W.I.$, that is:

$$\text{W.I.rate} = 100 \cdot \frac{\text{number of correct labels}}{500} \tag{2}$$

The results of the different methods are shown in Table II. One can see that most methods obtain a writer identification rate of about 65%. The best methods are PRIP02-combination and TUA03-SVMOAA, which indeed obtain very similar results (77% and 76.6% respectively). These results demonstrate that the identification of the writer in graphical documents (such as music scores) is still challenging, and more research must be done.

V. Conclusion

The first music scores competition held in ICDAR has shown to wake up the interest of researchers, with 8 participant methods in the staff removal competition, and another 8 participant methods in the writer identification competition. The staff removal methods submitted by the participants have obtained very good performance in front of severe distorted images, although it has also been shown that there is still room for improvement, especially concerning the detection of the staff lines. Concerning writer identification, the participants' results have shown that more research is required for dealing with the identification of graphical documents. We hope that the competition results on the CVC-MUSCIMA database will foster the research on handwritten music scores in the near future.

Acknowledgment

We would like to thank Joan Casals and all the musicians who contributed to the CVC-MUSCIMA dataset; and Dr. Christoph Dalitz for providing the code which generates the staff distortions. This work has been partially supported by the Spanish projects TIN2008-04998, TIN2009-14633-C03-03, and CONSOLIDER-INGENIO 2010 (CSD2007-00018) and 2011 FIB 01022.

References

[1] H. Miyao, M. Maruyama, *An online handwritten music symbol recognition system*, International Journal on Document Analysis and Recognition 9(1), pp.49-58, 2007.

[2] A. Fornés, J. Lladós, G. Sánchez, X. Otazu, H. Bunke, *A combination of features for symbol-independent writer identification in old music scores*, International Journal on Document Analysis and Recognition 13(4), pp.243-259, 2010.

[3] A. Rebelo, G. Capela, J. Cardoso, *Optical recognition of music symbols*, International Journal on Document Analysis and Recognition 13(1), pp.19-31, 2010.

[4] 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, in press (DOI: 10.1007/s10032-011-0168-2), 2011.

[5] J.S. Cardoso, A. Capela, A. Rebelo, C. Guedes and J.P. da Costa, *Staff detection with stable paths*, IEEE Trans. on Pattern Analysis and Machine Intelligence 31(6), pp.1134-1139, 2009.

[6] I. Fujinaga, *Staff Detection and Removal*, in Visual Perception of Music Notation, S. George, ed., pp. 1-39, Idea Group, 2004.

[7] C. Dalitz, M. Droettboom, B. Pranzas and I. Fujinaga, *A Comparative Study of Staff Removal Algorithms*, IEEE Trans. on Pattern Analysis and Machine Intelligence 30(5), pp. 753-766, 2008.

[8] S. Al-Ma'adeed, E. Mohammed and D. Al Kassis, *Writer identification using edge-based directional probability distribution features for Arabic words*, Int'l Conf. on Computer Systems and Applications (AICCSA), pp. 582-590, 2008.

[9] S. Al-Ma'adeed, A.-A. Al-Kurbi, A. Al-Muslih, R. Al-Qahtani and H. Al Kubisi, *Writer identification of Arabic handwriting documents using grapheme features*, Int'l Conf. on Computer Systems and Applications (AICCSA), pp.923-924, 2008.

[10] C. Djeddi and L. Souici-Meslati, *A texture based approach for Arabic Writer Identification and Verification*, Int'l Conf. on Machine and Web Intelligence (ICMWI), pp. 115-120, 2010.