The 2012 Music Scores Competitions: 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
{afornes,adutta,agordo,josep}@cvc.uab.es

Abstract. Since there has been a growing interest in the analysis of handwritten music scores, we have tried to foster this interest by proposing in ICDAR and GREC two different competitions: Staff removal and Writer identification. Both competitions have been tested on the CVC-MUSCIMA database of handwritten music score images. In the corresponding ICDAR publication, we have described the ground-truth, the evaluation metrics, the participants' methods and results. As a result of the discussions with attendees in ICDAR and GREC concerning our music competition, we decided to propose a new experiment for an extended competition. Thus, this paper is focused on this extended competition, describing the new set of images and analyzing the new results.

Keywords: competition; graphics recognition; music scores; writer identification; staff removal.

1 Introduction

In the last years, there has been a growing interest in the analysis of handwritten music scores [1–3]. 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 [4–6], since a good detection and removal of the staff lines will allow the correct segmentation of the musical symbols, and consequently, will ease the correct recognition and classification of the music symbols.

Concerning writer identification in music scores, some approaches have been proposed in the last decade [7–9]. It must be said that 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. 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 (International Conference on Document Analysis and Recognition) and GREC (International Workshop on Graphics Recognition) two different competitions: Staff removal and Writer Identification. Both competitions have been tested on the CVC-MUSCIMA [1] database [10]. The CVC-MUSCIMA database has been designed for musical scores analysis and recognition. It consists of 1,000 handwritten music score images, written by 50 different musicians. Each writer has transcribed exactly the same 20 music pages, using the same pen and paper.

Details on these two competitions (ground-truth, metrics, participants' methods and results) can be found in the corresponding ICDAR publication [11]. In this paper, however, we would like to focus on the extended competition on staff removal that has been organized after ICDAR and GREC. In fact, during ICDAR and GREC, we received interesting feedback from researchers in music analysis. One of the most common suggestions was to use combinations of distortions for further staff removal competitions. As a result, we decided to generate a new set of images and ask the participants on the staff removal task to participate.

The rest of the paper is organized as follows. Firstly, we will briefly describe the music scores competition proposed in ICDAR and GREC. The writer identification competition is described in Section 2 and 3 describes the staff removal competition. Afterwards, Section 4 is devoted to the extended staff removal competition, describing the new set of images and analyzing the results obtained by the participant's methods. Finally, Section 5 concludes the paper.

2 Writer Identification Competition

For the writer identification competition, the CVC-MUSCIMA dataset [10] 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, because they are particularly useful here: 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 (see Fig.1) make easy the participation of researchers that do not work on staff removal. The staff lines were initially removed using color cues and manually checked for correcting errors (see more details in [10]).

2.1 Participants

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

PRIP02 These methods were submitted by Abdelâali Hassaïne and Somaya Al-Ma'adeed from the Pattern Recognition and Image Processing Research Group

[1] http://www.cvc.uab.es/cvcmuscima

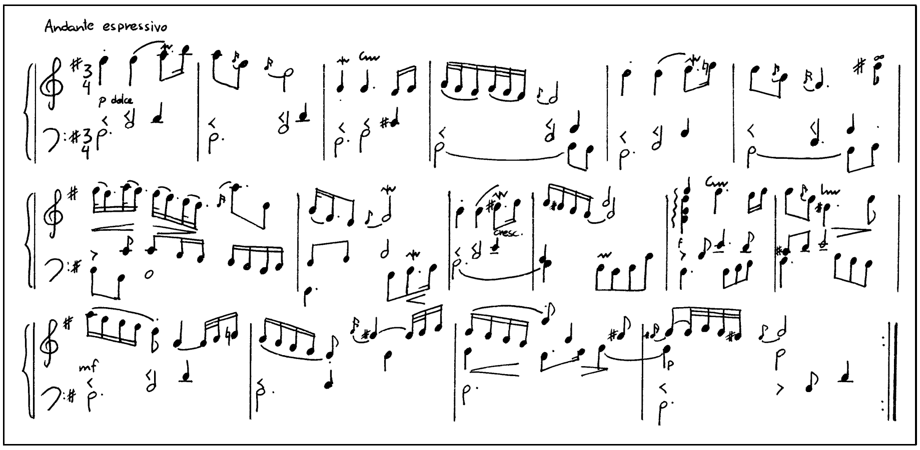


Fig. 1: Example of image without staff lines.

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 [13]).
- PRIP02-grapheme: The second one uses grapheme features, described in [14].
- 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.

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. They 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 [16].

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

- TUA03-5NN: A 5 nearest neighbor classifier with cityblock Distance Metric.

– TUA03-SVMOAO: Support Vector Machine (SVM One against one).
– TUA03-SVMOAA: Support Vector Machine (SVM One against all).
– TUA03-MLP: Multilayer perceptron (MLP).
– TUA03-Combination: A combination of the four previous classifiers. The combination rule used is Majority Vote.

2.2 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.*, which corresponds to the number of correctly identified documents divided by the total amount of documents:

$$\text{W.I.rate} = 100 \cdot \frac{\text{number of correctly identified documents}}{500} \tag{1}$$

The results of the different methods are shown in Table 1. One can see that most methods obtain a writer identification rate of about 65%. We can see that the best methods are PRIP02-combination and TUA03-SVMOAA, which indeed obtain very similar results (77% and 76.6% respectively). These results are obtained after the combination of different sets of features (PRIP02), or several classifiers (in case of TUA03).

Table 1: Writer Identification results. Number of correctly identified images and the final Writer Identification (W.I.) rate in %.

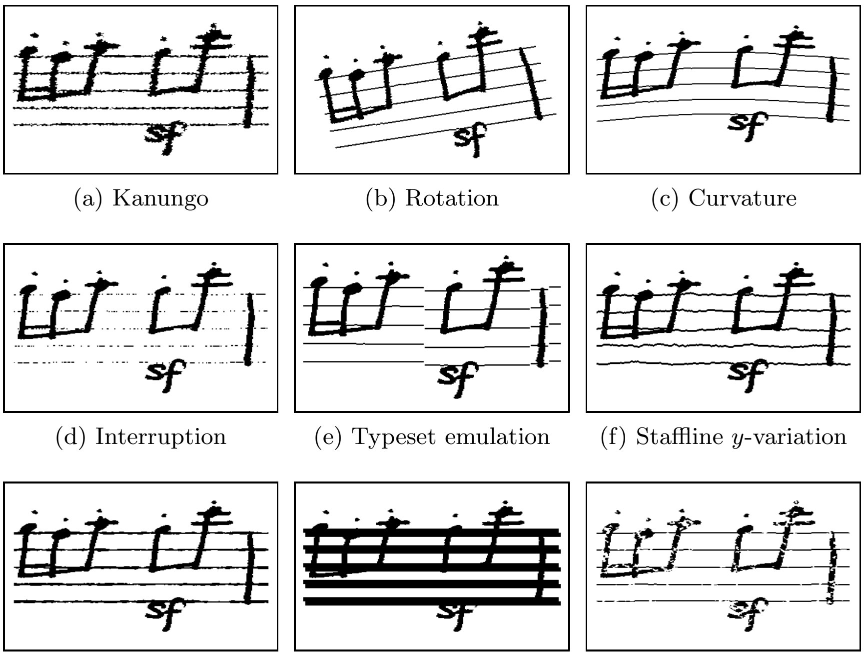
Method	Correct/Total	W.I.rate (%)
PRIP02-edges	327/500	65.4
PRIP02-grapheme	319/500	63.8
PRIP02-combination	385/500	**77.0**
TUA03-5NN	267/500	53.4
TUA03-MPL	324/500	64.8
TUA03-SVMOAA	383/500	**76.6**
TUA03-SVMOAO	333/500	66.6
TUA03-combination	352/500	70.4
Specific IJDAR	425/500	**85**

It must be said that in the CVCMUSCIMA publication [10], the reference writer identification rate is about 85%. These results are obtained using a specific writer identification method for music scores, which is based on the bag-of-notes approach described in [15]. In addition, the authors of [17] demonstrate that their specific method also obtains better results than some writer identification methods for roman text documents that are adapted for music scores.

Since the competition results reported here are obtained by adapting writer identification methods for arabic documents, we could conclude that specific methods might be the best choice. In summary, all 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.

3 Staff Removal Competition

For testing the robustness of the staff removal algorithms, we have applied the following distortion models (see Fig.2) 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 [10] for details.



(a) Kanungo (b) Rotation (c) Curvature

(d) Interruption (e) Typeset emulation (f) Staffline y-variation

(g) Staffline thickness-v (h) Staff thickness Ratio (i) White speckles

Fig. 2: Examples of Staff deformations.

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 6000 images will be used for testing them.

3.1 Participants

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

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.

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 [4]. 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 $whiterun = 2 \cdot staffspaceheight$ 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.

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, 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.

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. They 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 [12]. 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 [5] (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 [5] (Section 3.4). The source code is available in the website: *http://music-staves.sourceforge.net/* (class $MusicStaves\_skeleton$).

3.2 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:

$$\text{E.R.} = 100 \cdot \frac{\#\text{misclassified } sp + \#\text{misclassified non } sp}{\#\text{all } sp + \#\text{all non } sp} \tag{2}$$

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 participants' methods have been evaluated in two ways:

- Error rate without rejection: The average error rate is computed as the mean of the error rate of the images that the system could evaluate. Thus, the rejected images are not included here.
- Error rate with rejection: The average error rate is computed taking into account all the set of images (the 500 images of each kind of distortion). Thus, the rejected images are considered to have an E.R.=100%.

The results of the different methods are shown in Table 2. 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. Not surprisingly, most methods obtain very similar results when dealing with the original ideal images (1.5%) and Kanungo noise (2.85%). However, differences are significant when dealing with Curvature, Interruption or Thickness Ratio.

It is interesting to notice that there is no agreement in the kind of distortion that all staff removal methods solve in the best way. This means that some methods are more suitable to a specific kind of distortion, whereas others solve in a better way another kind of distortions.

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 must be said 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).

Table 2: Staff Removal results. Error Rate (E.R.) in % for each one of the 12 distortions. For each one of the participant's methods, we show the Error Rate with rejection (With R.) and without rejection (No R.). In case of the Error rate without rejection, we also show the number # of rejected images. The last row corresponds to the overall Error Rate.

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

4 Staff Removal Extended Competition

During the GREC 2011 workshop and the ICDAR 2011 conference, we had the opportunity to discuss about our music competition with some attendees and researchers working on music analysis. As a result of this feedback, we decided to contact the staff removal participants again in order to propose them to run their algorithms with a new set of distorted images.

The goal of this new experiment is to test the robustness of the participants' algorithms in a more realistic case: a combination of different distortions. The idea behind is that in a real scenario, a document may contain several kinds of distortions.

Next, we will describe the new set of images and the obtained results.

4.1 Images Description

In this new experiment, for each one of the original image, we have generated one distorted image. Thus, the new set of distorted images is composed of 1000 images, 500 images for training and 500 images for testing. These images have been generated by the combination of four different distortion methods: staffline y-variation, curvature, white speckles and kanungo. However, and contrary to the GREC and ICDAR competition, we have generated three different levels of distortions, namely low, medium and severe. Table 3 shows the number of images for each level of distortion and the parameters used for each distortion. The staff distortion code that has been used for generating these images is available in the CVC-MUSCIMA website.

Table 3: Staff Removal - Extended Competition. Number of images and parameters used.

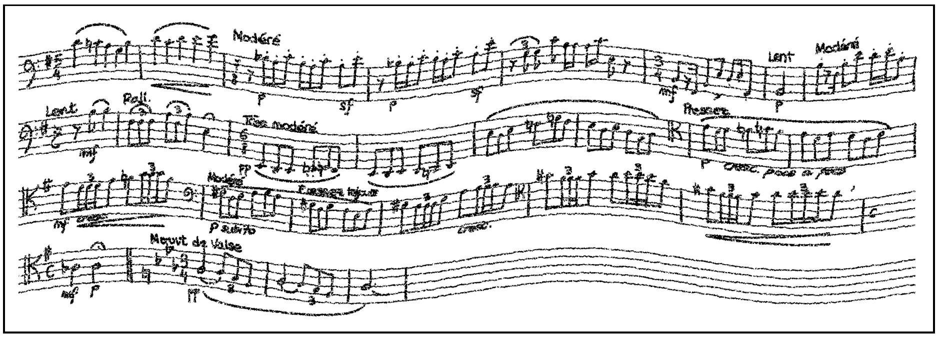
Distortion Level	Number of Images	staffline y-variation (maxdiff,c)	curvature (amplitude, period)	white speckles (p, n, k)	kanungo (eta, a0, a, b0, b, k)
Low	200	(1, 0.5)	(0.05, 1)	(0.03, 6, 2)	(0, 1, 1, 1, 1, 2)
Medium	200	(3, 0.7)	(0.05, 5)	(0.04, 8, 2)	(0 , 1 , 1, 1, 1, 2)
Severe	100	(5, 0.9)	(0.05, 8)	(0.05, 10, 2)	(0 , 1 , 1, 1, 1, 2)

As an illustrative example, Figure 3 shows the same music score with a low, medium and severe distortion. Notice that, although the original document is the same, the resulting images look quite different, especially in terms of curvature and noise (see Fig.(d-e)).

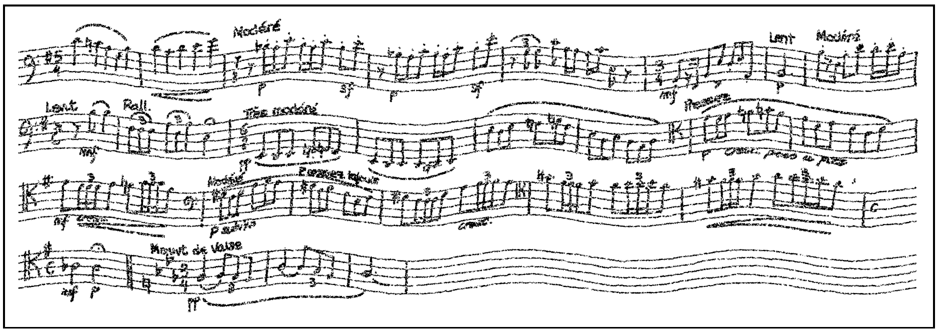
4.2 Results

The performance of the algorithms in this new set of images has been measured using the same metrics as the ones used for the previous staff removal competition: Error Rate (E.R.) with and without rejection. Table 4 shows the results of

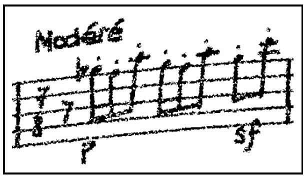
(a) Low distortion



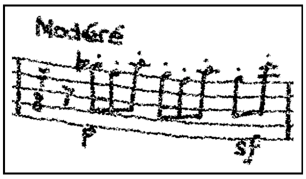
(b) Medium distortion



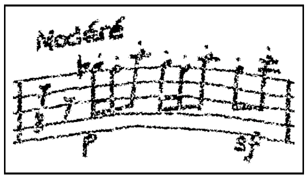
(c) Severe distortion



(d) Low distortion



(e) Medium distortion



(f) Severe distortion

Fig. 3: Combination of distortion methods applied to the same music score and with different distortion levels. (a) Low, (b) Medium and (c) Severe distorted images. Captions of these images: (d) Caption of the image (a), (e) Caption of the image (b), and (f) Caption of the image (c).

the extended staff removal competition. In most cases, the Error Rate increases when increasing the level of distortion (e.g. ISI01, NUG04-Fuji, etc.). However, some of the methods are quite stable, with a variation of less than one percent, such as INP02 (5.86-6.52%) or NUG04-LTr (11.2-10.6%).

Comparing with the results from the previous staff removal competition, there are two main aspects to remark. Firstly, and as expected, these error rates are higher than the ones shown in Table 2. The main reason is that these methods could be individually trained to cope with each isolated distortion, in other words, the parameters could be different depending on the distortion to be treated. However, when images are generated through a combination of distortions, it is hardly impossible to find one set of parameters that is suitable for all the distortions. Secondly, the amount of rejected images is lower than the rejected images in Table 2). This can be explained because, although the resulting image is more complex, the staff lines look more realistic and consequently, the staff detection method has a better performance.

One interesting aspect is that, in this extended competition, the best results are obtained by INP02, followed by NUG04-LTr (with an overall Error Rate of 6.19% and 10.63%. respectively). Consequently, these methods are a priori more suitable for dealing with realistic images (complex images with several kinds of distortions) than the best methods (ISI01 and NUS03) described in the previous staff removal competition, which seem to be very sensitive to parameter configuration.

Table 4: Staff Removal results of the Extended Competition. For each submitted method and distortion level, we show the Error Rate (E.R.) in % with rejection (With R.) and without rejection (No R.), and the number # of rejected images. The last row corresponds to the overall (average) Error Rate.

Distortion	Error	ISI01	INP02	NUS03	NUG04-Fuji	NUG04-LTr	NUG04-Skel
Low-	No R.	14.1	5.86	39.6	21.3	11.2	11.7
(200	#	0	0	2	5	0	3
Images)	With R.	14.1	5.86	40.6	23.8	11.2	13.2
Medium-	No R.	16.2	6.19	54.6	30.6	10.1	14
(200	#	0	0	0	18	2	4
Images)	With R.	16.2	6.19	54.6	39.6	11.1	16
Severe-	No R.	21.2	6.52	49.3	40.8	10.6	15.3
(100	#	0	0	0	0	0	4
Images)	With R.	21.2	6.52	49.30	40.8	10.6	19.3
Overall	No R.	**17.17**	**6.19**	**47.83**	**30.9**	**10.63**	**13.67**
Error	#	0	0	2	23	2	11
Rate	With R.	17.17	6.19	48.17	34.73	10.97	16.17

5 Conclusion

The first music scores competition held in ICDAR2011 and GREC2011 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.

In the writer identification task, the participants' results have shown that more research is required for dealing with the identification of graphical documents. In this context, and since the adaptation of writer identification methods from other kind of documents have obtained modest results, one may conclude that specific approaches for music scores are the best choice.

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. In addition, we have extended this competition by adding a new set of images which have been generated from a combination of different kinds of distortions. The new results of the participants have demonstrated that most methods significantly decrease their performance when dealing with a combination of distortions, which is precisely a more realistic scenario than the previous one. In future competitions, it would be also interesting to see how the systems could cope with real distortions, especially the ones that appear in historical documents.

Finally, we hope that the competition results on the CVC-MUSCIMA database will foster the research on handwritten music scores in the near future.

Acknowledgments. 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. We thank all the participants, because without their interest the competition would not have been possible. We thank the writer identification participants: Abdelâali Hassaïne, Somaya Al-Ma'adeed, Chawki Djeddi, Labiba Souici-Meslati. Especially, we would like to thank the researchers that kindly agreed to participate in the extended staff removal competition: Jit Ray Chowdhury, Dr. Umapada Pal, Ana Rebelo, Dr. Jaime S. Cardoso, Bolan Su, Dr. Shijian Lu, Dr. Chew-Lim Tan, Dr. Christoph Dalitz and Andreas Kitzig.

This work has been partially supported by the Spanish projects TIN2011-24631, TIN2009-14633-C03-03, 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. 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.
5. 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.
6. J. Cui, H. He Y. Wang, *An adaptive staff line removal in music score images*, IEEE 10th International Conference on Signal Processing (ICSP), pp.964-967, 2010.
7. , N. Luth, *Automatic identification of music notations*, Proceedings of the Second International Conference on WEB Delivering of Music (WEDELMUSIC), pp.203-210, 2002.
8. , I. Bruder, I. Temenushka, L. Milewski, *Integrating knowledge components for writer identification in a digital archive of historical music scores*, Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital libraries (JCDL), pp.397-397, 2004.
9. , S. Marinai, B. Miotti, G. Soda, *Bag of Characters and SOM Clustering for Script Recognition and Writer Identification*, International Conference on Pattern Recognition, pp.2182-2185, 2010.
10. 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.
11. A. Fornés, A. Dutta, A. Gordo and J. Lladós, *The ICDAR 2011 Music Scores Competition: Staff Removal and Writer Identification*, International Conference on Document Analysis and Recognition, pp.1511-1515, 2011.
12. I. Fujinaga, *Staff Detection and Removal*, in Visual Perception of Music Notation, S. George, ed., pp. 1-39, Idea Group, 2004.
13. S. Al-Ma'adeed, E. Mohammed and D. Al Kassis, *Writer identification using edge-based directional probability distribution features for Arabic words*, International Conference on Computer Systems and Applications (AICCSA), pp. 582-590, 2008.
14. 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.
15. A. Gordo, A. Fornés, E. Valveny, J. Lladós, *A bag of notes approach to writer identication in old handwritten musical scores*, International Workshop on Document Analysis Systems, pp. 247-254, 2010.
16. 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.
17. A. Fornés, J. Lladós, *A Symbol-Dependent Writer Identification Approach in Old Handwritten Music Scores*, International Conference on Frontiers in Handwriting Recognition, pp. 634 - 639, 2010.