International Journal on Document Analysis and Recognition manuscript No.
(will be inserted by the editor)

CVC-MUSCIMA: A Ground-Truth of Handwritten Music Score Images for Writer Identification and Staff Removal

Alicia Fornés · Anjan Dutta · Albert Gordo · Josep Lladós

Received: date / Accepted: date

Abstract The analysis of music scores has been an active research field in the last decades. However, there are no publicly available databases of handwritten music scores for the research community. In this paper we present the CVC-MUSCIMA database and ground-truth of handwritten music score images. The dataset consists of 1,000 music sheets written by 50 different musicians. It has been especially designed for writer identification and staff removal tasks. In addition to the description of the dataset, ground-truth, partitioning and evaluation metrics, we also provide some baseline results for easing the comparison between different approaches.

Keywords Music Scores · Handwritten Documents · Writer Identification · Staff Removal · Performance Evaluation · Graphics Recognition · Ground-truths

1 Introduction

The analysis of music scores [19,22,31,35] is a classical area of interest of Document Image Analysis and Recognition (DIAR). Traditionally, the main focus of interest within the research community has been the transcription of printed music scores. Optical Music Recognition (OMR) [1,2,15] consists in the understanding of information from digitized music scores and its conversion into a machine readable format. It allows a wide variety of applications such as the edition of scores never edited, renewal of old scores, conversion of scores into Braille, production of audio files, adaptation of existing works to other instrumentations, transposing a music sample to some other clef or key signature, producing parts from a given score or a full score from given parts, creation of collecting databases to perform musicological analysis. Since the first works by Prerau and Pruslin in the late 1960's [27,26], interest in OMR has grown in last decades, appearing several complete OMR systems for printed music (such as Aruspix, Gamera or Guido [28,29,33]), braille music approaches [3], and even an almost real-time keyboard-playing robot (the Wabot-2 robot [21]).

Among the required stages of an Optical Music Recognition system, an special emphasis has been put in the staff removal algorithms [5,6,12,32], 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 detection, recognition and classification of the music symbols. Staff removal is somehow related to form processing [18], where ruling lines must be removed prior to recognize the text. The main difference is that staff removal techniques can take advantage of grouping rules, in other words, the algorithm can search a group of five equidistant horizontal lines (the staff).

In the last decade, there has been a growing interest in the analysis of handwritten music scores [11,20, 23,24,30,31,34]. In this context, the focus of interest is two-fold: the recognition of handwritten music scores, and the identification (or verification) of the authorship of a music score. Concerning writer identification, musicologists do not only perform a musicological analysis of the composition (melody, harmony, rhythm, *etc.*), but also analyse the handwriting style of the manuscript. In this sense, writer identification can be performed by

A. Fornés, A. Dutta, A. Gordo, J. Lladós
Computer Vision Center - Dept. of Computer Science
Universitat Autònoma de Barcelona
Edifici O, 08193, Bellaterra, Spain
Tel.: +34-935811828
Fax: +34-935811670
E-mail: {afornes,adutta,agordo,josep}@cvc.uab.es

analyzing the shape of the hand-drawn music symbols (*e.g.* music notes, clefs, accidentals, rests, *etc.*), because it has been shown (see [10]) that the author's handwriting style that characterizes a piece of text is also present in a graphic document. Nevertheless, musicologists must work very hard to identify the writer of a music score, especially when there is a large amount of writers to compare with. Recently, several writer identification approaches have been developed for helping musicologists in such a time consuming task. These approaches are based in many different methodologies, such as Self Organizing Maps [20], Bag of Features [14], knowledge-based approaches [4,13], or even systems which adapt some writer identification approaches for text documents to music scores [11].

Contrary to printed music scores databases [6], there are no public databases of handwritten music scores available for the research community. For this reason, there is a need of a public database and ground-truth for validating the different methodologies developed in this research field. With this motivation, in this paper we present the CVC-MUSCIMA [1] ground-truth: a ground-truth of handwritten music score images. The database and ground-truth are available in the website: http://www.cvc.uab.es/cvcmuscima. This dataset consists of 1,000 music sheets written by 50 different musicians, and has been especially designed for writer identification and staff removal tasks.

In this paper we describe the database, evaluation metrics, partitions (data subsets) and baseline results for comparison purposes. We believe that the presented ground-truth will serve as a basis for research in handwritten music analysis. Moreover, we will show that the effort of generating ground-truth can be reduced by using color cues and by applying distortions to both original images and ground-truth images.

The rest of the paper is organized as follows. Section 2 describes the dataset and the staff distortions applied. Section 3 presents the evaluation partitions, metrics and baseline results for comparison purposes. Finally, concluding remarks are described in Section 4.

2 Dataset

The dataset consists of 20 music pages of different compositions transcribed by 50 writers, yielding a total of 1,000 music pages. All the 50 writers are adult musicians (ages from 18 to 35) in order to ensure that they have their own characteristic handwriting style. We chose the set of 50 musicians as much heterogeneous as possible. The musicians are from different geographic locations (different cities in Spain). The set of writers includes advanced musician students (in conservatories of music or at University), musicologists, music teachers and professional musicians, but as far as we know, none of them are famous. They all have been studying music for many years, and consequently, they have their own characteristic handwriting style. Figure 1 shows some examples of handwritten music scores written by three different musicians. Having a look at the images, one can see that writer B tends to write in a rectilinear way (with very thin headnotes), while writers A and C draw very round headnotes. In addition, it can be observed that writer C tends to write short symbols (and also short slurs), whereas writers A and B draw taller music symbols and longer slurs.

Each writer has been asked to transcribe exactly the same 20 music pages, using the same pen (a black Pilot v7 Hi-Tecpoint) and the same kind of music paper (standard DIN A4 sheets with printed staff lines in blue color). The set of the 20 selected music sheets contains monophonic and polyphonic music, and it consists of music scores for solo instruments (*e.g.* violin, flute, violoncello or piano) and music scores for choir and orchestra. It must be noted that the music scores only contain the handwriting text considered as part of the music notation theory (such as dynamics and tempo notation), and for this reason, music scores for choir do not contain lyrics.

Furthermore, for staff removal tasks, each music page has been distorted using different transformation techniques (please refer to Section 2.2 for details), which, together with the originals, yields a grand total of 12,000 base images.

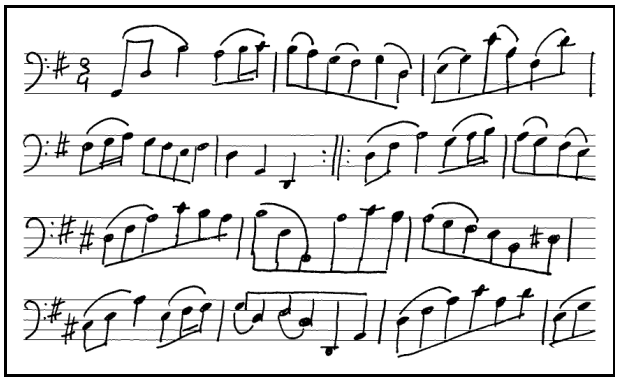
Next, we describe the data acquisition, the generated deformations and the different ground-truths and data formats.

2.1 Acquisition and Preprocessing

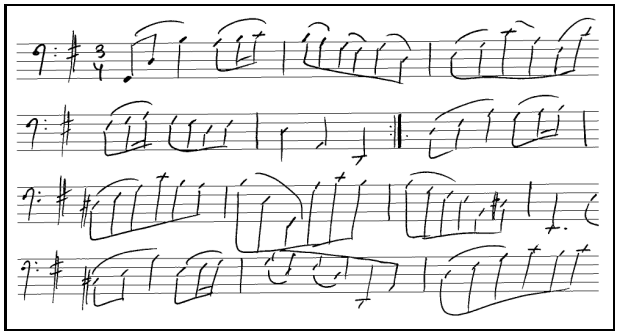
Documents were scanned using an flatbed Epson GT-3000 scanner set at 300 dpi and 24 bpp, as colour cues were used in the original templates to ease the elaboration of the staff ground-truth. Later, the images were converted to 8 bit gray scale. Care was put into obtaining a good orientation during the scanning stage, and absolutely no digital skew correction was applied once the pages were scanned.

The staff lines were initially removed using color cues. Afterwards, they were binarized and manually checked for correcting errors, specially when some segments of the staff lines were manually added by the

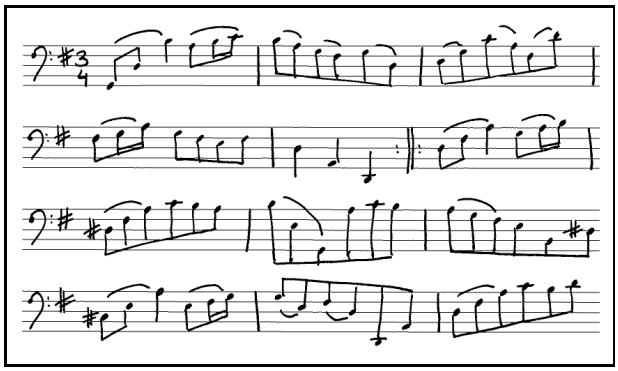
[1] CVC-MUSCIMA stands for **C**omputer **V**ision **C**enter - **MU**sic **SC**ore **IMA**ges



(a) Writer A



(b) Writer B



(c) Writer C

Fig. 1: Examples of pieces of music scores written by three different musicians. Notice the differences in handwriting styles.

writer (see an example in Fig. 2). Thus, from the gray scale images, we generated the binarized images, the images with only the music symbols (without staff lines), and finally, the images with only the staff lines. Next, we describe the distortions applied to the music scores for staff removal.

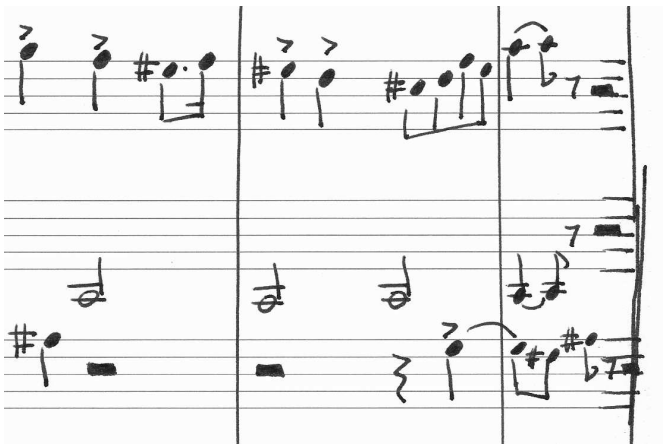


Fig. 2: Example of a section of a music score with some segments of hand-drawn staff lines

2.2 Staff Distortions

To test the robustness of different staff removal algorithms, we have applied a set of distortion models to our music score images. These distortion models are inspired by the work of Dalitz et al. [6] for testing the performance of staff removal algorithms in printed music scores. In [6] the authors describe nine different types of deformations for simulating their dataset with real world situation: Degradation with Kanungo noise, Rotation, Curvature, Staffline interruption, Typeset emulation, Staffline y-variation, Staffline thickness ratio, Staffline thickness variation and White speckles.

In order to obtain the same effect, the deformation is simultaneously applied to the original and the ground-truth staff images, which correspond to binary images with only the staff lines. A brief description of the individual deformation models is given next:

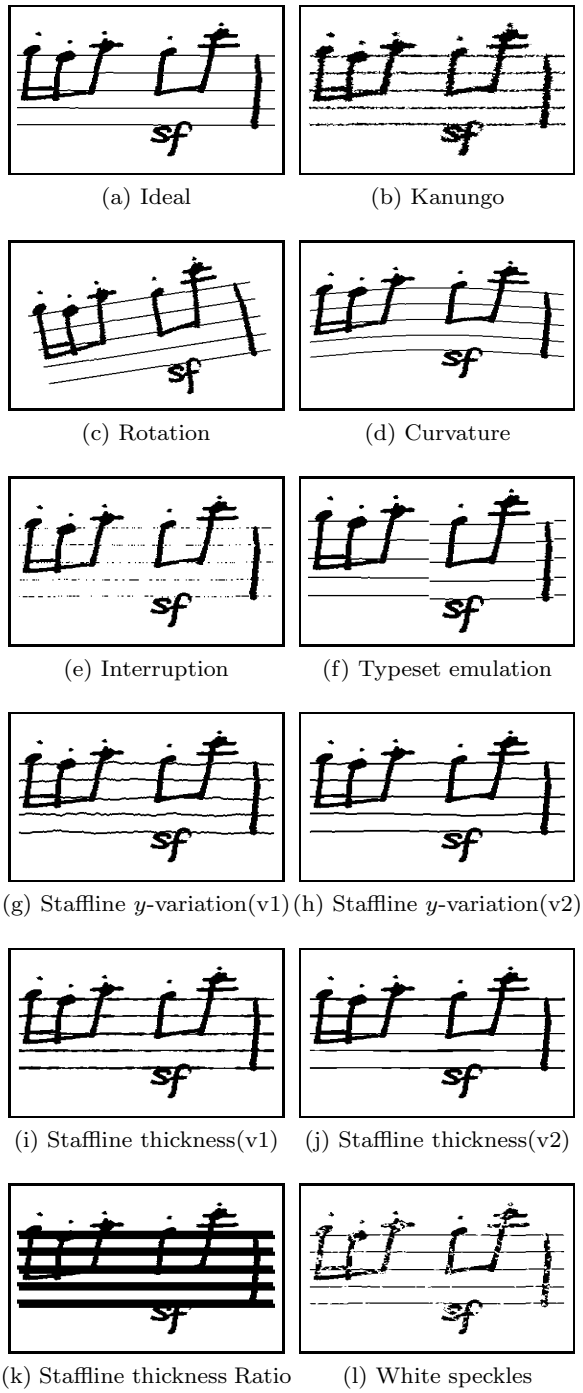
- *Kanungo noise.* Kanungo et al [17] have proposed a noise model to create local distortions introduced during scanning. The model mainly affects to the contour pixels and has little effect on the inner pixels (see Fig. 3b).
- *Rotation.* The distortion rotation (see Fig. 3c) consists in rotating the entire staff image by the specified parameter angle.
- *Curvature.* The curvature is performed by applying a half sinusoidal wave over the entire staffwidth. The strength of the curvature is regulated by a parameter which is a ratio of the amplitude to the staffwidth (see Fig. 3d).
- *Staffline interruptions.* The staffline interruptions consist in generating random interruptions with random size in the stafflines. This model mainly affects to the staffline pixels, and simulates the scores that

are written on already degraded stafflines (see Fig. 3e).
- *Typeset emulation*. This particular defect is intended to imitate the sixteenth century prints that are set by lead types. Consequently, they have staffline interruptions between symbols and also a random vertical shift of each vertical staff slice containing a symbol (see Fig. 3f).
- *Staffline y-variation and Staffline thickness variation*. These kind of defects are created by generating a Markov chain describing the evolution of the y-position or the thickness from left to right. This is done since, generally the y-position and the staff thickness values for a particular x-position depend on its previous x-position (Fig. 3g, 3h, 3i, and 3j show some examples of these deformations with different parameters).
- *Staffline thickness ratio*. This defect only affects to the whole staffline thickness of the music score, which consists in generating stafflines of different thickness (see Fig. 3k).
- *White speckles*. This degradation model is used to generate white noise within the staff pixels and musical symbols (see Fig. 3l).

Table 1 describes the parameters of the respective models. Dalitz et al. [6] have developed the MusicStaves toolkit [2], which is available for reproducing the experiments in other datasets. However, these available algorithms for distorting the staff lines have an important drawback: they require computer generated perfect artificial images, which means perfect horizontal staff lines, equidistant, and also with the same thickness. Since our dataset contains printed and handwritten segments of staff lines (see Fig. 2), their algorithms can not be directly applied to our music scores. For this reason, we have modified these algorithms to reproduce the same distortion model in our handwritten music scores (where we do not assume any constraints for perfect staff lines).

For validating the staff removal algorithms, we have generated a set of 11,000 distorted images by applying the nine already described distortion models, where two of them have been applied twice (see Fig.3). Thus, for each original image, we have obtained 11 distorted images by applying these distortion algorithms with the parameters described in Figure 3. As a result, the dataset for staff removal purposes contains 12,000 images (1,000 original images plus the 11,000 distorted ones). However, since we also provide the code of the

[2] http://lionel.kr.hs-niederrhein.de/~dalitz/data/projekte/stafflines/doc/musicstaves.html



(a) Ideal (b) Kanungo

(c) Rotation (d) Curvature

(e) Interruption (f) Typeset emulation

(g) Staffline y-variation(v1) (h) Staffline y-variation(v2)

(i) Staffline thickness(v1) (j) Staffline thickness(v2)

(k) Staffline thickness Ratio (l) White speckles

Fig. 3: Staff deformation and their corresponding parameters. (a) Ideal image. (b) Kanungo $(\eta, \alpha_0, \alpha, \beta_0, \beta, k) = (0, 1, 1, 1, 1, 2)$. (c) Rotation$(\theta) = (12.5°)$. (d) Curvature $(a, p) = (0.05, 1.0)$. (e) Staffline interruptions$(\alpha, n, p) = (0.5, 3, 0.5)$. (f) Typeset emulation $(n, p, ns) = (1, 0.5, 10)$. (g)-(h) Staffline y-variation $(n, c) = (5, 0.6)$ and $(n, c) = (5, 0.93)$. (i)-(j) Staffline thickness variation $(n, c) = (6, 0.5)$ and $(n, c) = (6, 0.93)$. (k) Staffline thickness ratio $(r) = (1.0)$. (l) White speckles $(p, n, k) = (0.025, 10, 2)$.

Table 1: Image deformation and corresponding parameters. For more information about the parameters and the generation of each distortion, refer to [6].

Deformation	Parameters description
Kanungo noise $(\eta, \alpha_0, \alpha, \beta_0, \beta, k)$	Each foreground pixel is flipped with probability $\alpha_0 e^{-\alpha d^2+\eta}$ (d is the distance to the closest background pixel); each background pixel is flipped with probability $\beta_0 e^{-\beta d^2+\eta}$ (d is the distance to the closest foreground pixel).
Rotation (θ)	θ is the rotation angle to be applied.
Curvature (a, p)	a is the amplitude of the sine wave divided by the staffwidth, p is the number of times a half sine wave should appear in the entire staff width.
Staffline interruption (α, n, p)	α is the probability for each pixel to be the center of an interruption, n and p are the parameters for the binomial distribution of size interruption.
Typeset emulation (n, p, ns)	n and p are the parameters for the binomial distribution deciding the horizontal gaps, ns is the parameter for another binomial distribution deciding the y gap where the value of the other parameter is always 0.5.
Staffline y-variation (n, c)	n and $p = 0.5$ are the parameters of the binomial distribution deciding the stationary distribution of Markov chain, c is an inertia factor allowing the smooth transition.
Staffline thickness ratio (r)	r is the ratio of the $staffline\_height$ to $staffspace\_height$.
Staffline thickness variation (n, c)	n and $p = 0.5$ are the parameters of the binomial distribution deciding the stationary distribution of Markov chain, c is an inertia factor allowing the smooth transition.
White speckles (p, n, k)	p is the parameter for speckle frequency, n is the size of the speckle and k is the size of the structuring element used for closing operation.

staff distortions algorithms, the users can generate the distorted images with their desired parameters.

3 Ground-truth

In this section we describe the images, evaluation partitions (subsets), evaluation metrics and some baseline results. Thus, they will serve as a benchmark scenario for a fair comparison between different approaches.

Concerning the baseline results, it must be said that since the main contribution of this work is the framework for performance evaluation, we include some baseline results just for reference purposes.

Table 2: Image flavours designed for writer identification and staff removal tasks. Recommended images for each task in bold.

Task	Images provided
Writer Ident.	1,000 original undistorted grey scale images 1,000 binary images (with staff lines) 1,000 **binary staffless images**
Staff Removal	12,000 **binary images with staff lines** 12,000 binary images of *only* staff lines 12,000 **binary staffless images**

3.1 Images Description

All the images of the dataset are presented in PNG format. Each document of the dataset (1,000 original images plus the 11,000 distorted images) is labelled with its writer identification code and presented in different image flavours:

- Original grey scale image (only for the original 1,000 images).
- Binary image (with staff lines).
- Binary staffless image (only music symbols).
- Binary staff lines image (no music symbols).

Although all this information is available for all tasks, we encourage the use of certain image flavours for different tasks. The staffless images 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 dependant on the performance of the particular staff removal technique applied (see an example in Figure 4).

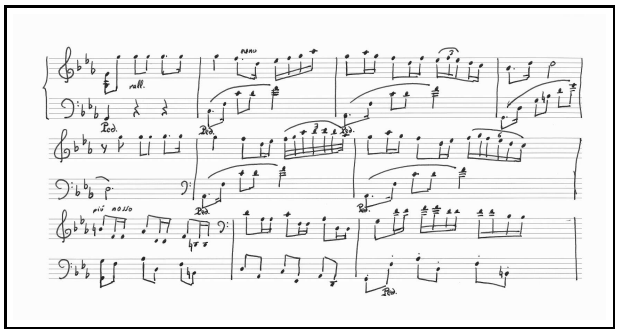
Similarly, for the staff removal tasks, staff lines images without music symbols (see Fig. 5) may be useful, not only for the evaluation of the method but also for training purposes. It must be said that the ground-truth images only show the pixels that belong only to staff lines. Consequently, these images contain holes, which correspond to the pixels belonging to music symbols.

Table 2 summarizes the provided images and these recommendations.

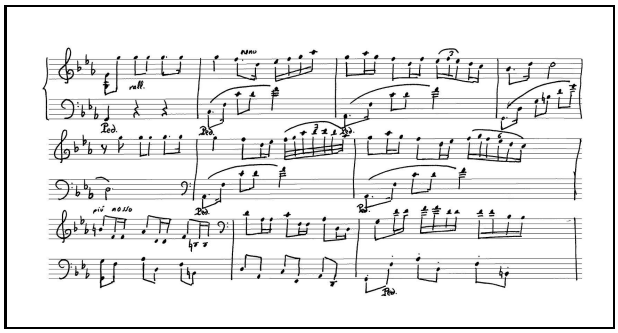
3.2 Evaluation Partitions for Writer Identification

For training and evaluation purposes, we devised two sets of ten partitions, which were especially designed for the evaluation of writer identification tasks:

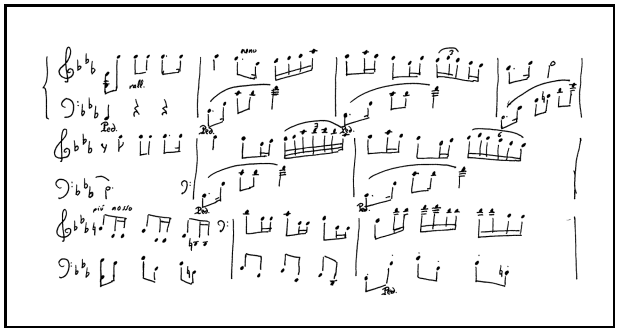
Set A, or constrained. In the first set of partitions, the training pieces of a given fold are the same for each writer, and so none of the pieces of the test set



(a) Gray image

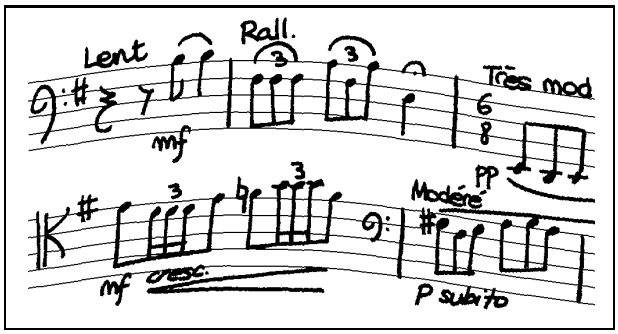


(b) Binary image

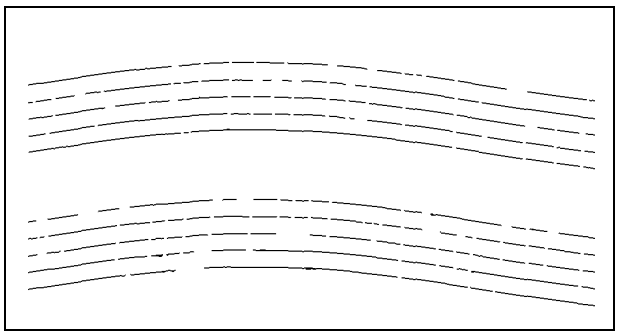


(c) Image without staff lines

Fig. 4: Example of the ground-truth of music scores for writer identification: (a) Gray image, (b) Binary image, (c) Staffless binary image



(a) Curved Image



(b) Staff-only Curved Image

Fig. 5: Example of the ground-truth for staff removal: (a) Curved image, (b) Staff-only curved image. Notice that the staff-only image contains holes corresponding to pixels that belong to music symbols.

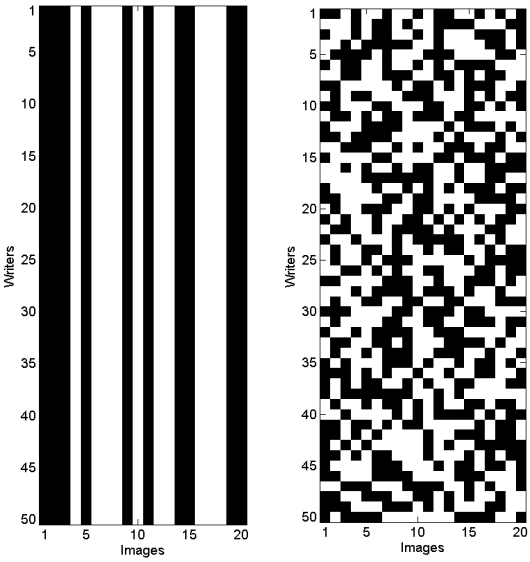
have been used during the training stage. As an illustrative example, look at Figure 6a. Since the first music page of one writer is in the train set of a given fold, all the first music pages of the remaining writers will also be in the train set of that particular fold.

Set B, or unconstrained. In the second set of partitions this constraint is not satisfied, and pieces that appear in the training set of one author will appear in the test set of a different one (for example, the first music page will appear in the train set of one author and in the test set of another, as seen in Figure 6b).

These partitions are particularly devised to attest that we are indeed performing writer identification instead of rhythm classification. Indeed, if the method was performing rhythm classification, it is reasonable to think that, in set B, unconstrained, test pieces from one author would be matched with the exact same pieces that appear in the train set of a different author, and so the classification results would be significantly lower than on set A, where this confusion is not possible. At the same time, a writer identification rate in set B similar to the one in set A will show that the system is classifying according to the handwriting style and not being particularly affected by the kind of music notes and symbols appearing in the music sheet.

In each partition, 50% of the documents of each writer belong to the training set and the other 50% belong to the test set. Furthermore, effort has been put in guaranteeing that each piece appears approximately 50% of the time in training and 50% in test. The exact partitions can be found in the dataset pack.

It must be said that instead of the proposed partitions, other strategies (such as "Leave-one-out") can be also used. However, we encourage the use of these partitions to test whether the system is rhythm dependant or not. In any case (partitions or "Leave-one-out"),



(a) Constrained set (b) Unconstrained set

Fig. 6: Train (black) and test (white) documents of each one of the 50 writers in a given fold in the constrained (left) and unconstrained (right) sets. In the constrained sets, all the writers use the same pieces for training. In the unconstrained sets, pieces used for training by one writer may appear in test for another writer.

the metrics described in next subsection can be applied without any modification.

3.3 Evaluation Metrics and Baseline Results for Writer Identification

In this subsection, we will describe the evaluation metrics and, as an illustrative example, some baseline results for writer identification purposes.

Metrics. Writer identification systems are evaluated considering two options: if the image has been correctly classified taking into account the n-first authors, or only the first writer. In our scenario, we will treat it as a binary problem, in which a music score is correctly classified only if the first nearest writer corresponds to the ground-truthed one.

Method. As we have commented, it is out of the scope of this work to make a comparison of different writer identification methods in the literature. However, and only for reference purposes, we provide baseline results using a recent writer identification method for musical scores [14].

In the Bag-of-Notes approach described in [14], features are computed using the Blurred Shape Model descriptor [8]. As in the Bag-of-Visual-Words framework, a codebook is built and symbols are assigned to the vocabulary words to represent the musical scores. Finally, they are classified using a SVM trained in a 1 *vs.* all fashion.

In that work, the authors presented a vanilla Bag-of-Notes, with the following properties:

- Unsupervised clustering with k-means.
- Hard assignment.
- Linear kernel.

Afterwards, they proposed the following modifications:

- Supervised clustering (learning a different codebook for each author and then merging the codebooks).
- Probabilistic vocabulary (learning the vocabulary with a GMM).
- Using a RBF kernel.

Interestingly, we found that the simpler vanilla implementation of the Bag-of-Notes obtained very similar results than the more complex modifications. In general, the probabilistic vocabularies bring little or no improvement over k-means even when tied with supervised clustering unless some adaptation is performed [25]; besides, the increasing size of the vocabulary usually makes these approaches impractical or unfeasible as the number of classes increase.

The RBF kernel provided slightly better results than the linear one. However, RBF has an extra parameter, the bandwitdh γ, which has to be validated. Also, using a linear kernel allows us to use solvers optimized for linear problems such as LIBLINEAR [9], which makes use of the cutting-plane algorithm and drastically improves the training speed of the SVM. To set the C trade-off cost of the SVM classifier, we used the same heuristic used by the SVM$^{\text{light}}$ [16] suite. Given a set of N training vectors $X = \{x_1, x_2, \ldots, x_N\}$, we set C as follows:

$$C = 1/k^2, \quad k = \frac{1}{N}\sum_{i=1}^{N} x_i x_i'. \tag{1}$$

This heuristic gave excellent classifications results, better than to those obtained by manually setting the parameter.

Because of its simplicity as well as comparatively good performance, we will report results using the vanilla implementation of the Bag-of-Notes (unsupervised clustering with k-means, hard assignment, and linear kernel), without the improvements of [14].

Results. Table 3 reports mean classification accuracy and standard deviation as a function of the number

Table 3: Mean classification accuracy (in %) and standard deviation as a function of the number of vocabulary words.

N. of words	Set A (*const.*)	Set B (*unconst.*)
16	31.60 ± 4.28	31.82 ± 2.06
32	43.18 ± 3.73	45.36 ± 3.36
64	57.08 ± 4.15	59.20 ± 2.83
128	73.02 ± 3.83	75.00 ± 2.06
256	84.72 ± 3.20	86.12 ± 1.35

of words for the two sets of partitions (please c.f. Section 3.2 for details on these partitions). Note that the accuracy results on both sets are quite similar, with a slight advantage for the second set; the higher accuracy and smaller standard deviation are probably caused because this set contains more variety in the training data. The fact that both sets obtain very similar results suggests that the Bag-of-Notes method is indeed performing writer identification and not rhythm identification, as would be the case if the *constrained* set obtained significantly better results than the *unconstrained* set.

3.4 Evaluation Metrics and Baseline Results for Staff Removal

The goal of staff removal is to delete those pixels that only belong to staff lines. If a pixel belongs to both a staff line and a musical symbol, then the pixel is labelled as belonging to the symbol. Consequently, the staff removal algorithm must be careful when removing the staff line segments, because they should not remove those pixels belonging to music symbols. Next we describe the evaluation metrics and some baseline results for staff removal purposes.

Metrics. Different metrics have been used in the literature. For example, in [6] the authors use the Error Rate, Segmentation Error and Staffline Interruptions, whereas in [32], the authors propose to use the percentage of staff lines falsely detected and the percentage of staff lines missed to detect. However, some of these measures are not very easy to compute. For this reason, we have chosen the pixel based evaluation metric to get the quantitative measurement of the performance of staff removal algorithms. These measures are very well-known, easy, efficient and fast to compute. In this scenario, we consider the staff removal problem as a two-class classification problem at the pixel level. For each of the images we compute the number of true positive pixels tp (pixels correctly classified as staff lines), false positive pixels fp (pixels wrongly classified as staff lines) and false negative fn (pixels wrongly classified as non-staff lines) by overlapping with the corresponding ground truth images. The Precision and Recognition Rate measures of the classification are computed as:

Table 4: Performance of the staff removal algorithm described in [7] (P = Precission, R = Recognition Rate and E = Error Rate are shown in %).

Deformation Type	P (%)	R (%)	E (%)
Ideal	99.22	93.56	2.9
Curvature	97.50	91.99	4.0
Interrupted	97.73	94.53	1.5
Kanungo	99.38	91.51	3.9
Rotated	96.71	93.38	3.7
Line Thickness Variation (v1)	95.45	94.96	5.1
Line Thickness Variation (v2)	97.45	93.97	4.5
Line y-Variation (v1)	94.54	93.63	6.7
Line y-Variation (v2)	94.73	94.33	6.3
Thickness Ratio	97.87	91.64	8.4
White Speckles	97.95	96.88	2.1
Typeset Emulation	98.86	89.46	4.6

$$\text{Precision} = P = \frac{tp}{tp + fp}, \tag{2}$$

$$\text{Recognition Rate} = R = \frac{tp}{tp + fn}. \tag{3}$$

The third metric error rate E *is computed as (#* means *"number of",* sp *means "staff pixels"):*

$$\text{E} = \frac{\#\text{misclassified } sp + \#\text{misclassified non } sp}{\#\text{all } sp + \#\text{all non } sp}. \tag{4}$$

Method. For the sake of illustration, we have chosen one of our staff removal algorithms as the baseline results. The approach proposed in [7] is based on the criteria of neighbouring staff components. It considers a staffline segment as a horizontal linkage of vertical black runs with uniform height, and then it uses the neighbouring properties of a staffline segment to discard the false segments.

Results. Table 4 shows the results of the staff removal algorithm using the proposed evaluation metrics and applied to the 12,000 distorted images. It must be noted that it does not obtain the best results in all cases with respect to the three evaluation metrics, showing that there is still room for research in this field. It should also be noted that these results are over the whole dataset and not only on the testing set, since this method does not require any training step.

4 Conclusions

In this paper we have described the CVC-MUSCIMA database and ground-truth, which has been especially designed for writer identification and staff removal tasks. We have also described the evaluation metrics, partitions and baseline results in order to ease the comparison between the different approaches that may be developed. It must be said that the main contribution of this work is the framework for performance evaluation, and for this reason, we have included some baseline results just for reference purposes. Concerning ground-truthing, we have shown that, although ground-truth generation is a time consuming task (specially when it is manually generated), one can reduce the effort of ground-truthing by using some simple methods (e.g. using color cues, applying distortions to images, and carrying the ground truth through to the distorted images).

The database can serve as a basis for research in music analysis. The database and ground-truth is considered complete at the current stage. However, further work will be focused on labelling each music note and symbol of the music score images for Optical Music Recognition purposes.

Acknowledgements We would like to thank all the musicians who contributed to the database presented in this paper. We would also specially thank to Joan Casals from the Universitat Autònoma de Barcelona for contacting with musicians, and collecting the music sheets. We would also like to thank 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. Bainbridge, D., Bell, T.: The challenge of optical music recognition. Computers and the Humanities **35**(2), 95–121 (2001)
2. Blostein, D., Baird, H.S.: Structured Document Image Analysis, chap. A critical survey of music image analysis, pp. 405–434. Springer Verlag (1992)
3. Bortolazzi, E., Baptiste-Jessel, N., Bertoni, G.: Bmml: A mark-up language for braille music. In: K. Miesenberger, J. Klaus, W. Zagler, A. Karshmer (eds.) Computers Helping People with Special Needs, *Lecture Notes in Computer Science*, vol. 5105, pp. 310–317. Springer Berlin / Heidelberg (2008)
4. Bruder, I., Ignatova, T., Milewski, L.: Knowledge-based scribe recognition in historical music archives. In: R. Heery, L. Lyon (eds.) Research and Advanced Technology for Digital Libraries, *Lecture Notes in Computer Science*, vol. 3232, pp. 304–316. Springer Berlin / Heidelberg (2004)
5. Cui, J., He, H., Wang, Y.: An adaptive staff line removal in music score images. In: Signal Processing (ICSP), IEEE 10th International Conference on, pp. 964–967. IEEE (2010)
6. Dalitz, C., Droettboom, M., Pranzas, B., Fujinaga, I.: A comparative study of staff removal algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(5), 753–766 (2008)
7. Dutta, A., Pal, U., Fornes, A., Llados, J.: An efficient staff removal approach from printed musical documents. Pattern Recognition, International Conference on pp. 1965–1968 (2010)
8. Escalera, S., Fornés, A., Pujol, O., Radeva, P., Sánchez, G., Lladós, J.: Blurred Shape Model for binary and grey-level symbol recognition. Pattern Recognition Letters **30**(15), 1424–1433 (2009)
9. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: Liblinear: A library for large linear classification. Journal of Machine Learning Research **9**, 1871–1874 (2008). Software available at http://www.csie.ntu.edu.tw/ cjlin/liblinear/
10. Fornes, A., Llados, J.: A symbol-dependent writer identification approach in old handwritten music scores. Frontiers in Handwriting Recognition, International Conference on pp. 634–639 (2010)
11. Fornés, A., Lladós, J., Sánchez, G., Otazu, X., Bunke, H.: A combination of features for symbol-independent writer identification in old music scores. International Journal on Document Analysis and Recognition **13**, 243–259 (2010)
12. Fujinaga, I.: Staff detection and removal. In: S. George (ed.) Visual Perception of Music Notation, pp. 1–39. Idea Group (2004)
13. Göcke, R.: Building a system for writer identification on handwritten music scores. In: Proceedings of the IASTED International Conference on Signal Processing, Pattern Recognition, and Applications (SPPRA), pp. 250–255. Rhodes, Greece (2003)
14. Gordo, A., Fornés, A., Valveny, E., Lladós, J.: A bag of notes approach to writer identification in old handwritten musical scores. In: Proceedings of the 9th IAPR International Workshop on Document Analysis Systems, DAS '10, pp. 247–254. ACM, New York, NY, USA (2010)
15. Homenda, W.: Computer Recognition Systems, chap. Optical music recognition: the case study of pattern recognition, pp. 835–842. Springer (2005)
16. Joachims, T.: Making Large-Scale Support Vector Machine Learning Practical. Advances in Kernel Methods. MIT-Press (1999). Software available at http://svmlight.joachims.org/
17. Kanungo, T., Haralick, R., Baird, H., Stuezle, W., Madigan, D.: A statistical, nonparametric methodology for document degradation model validation. IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(11), 1209 – 1223 (2000)
18. Lopresti, D., Kavallieratou, E.: Ruling line removal in handwritten page images. In: International Conference on Pattern Recognition, pp. 2704–2707. IEEE (2010)
19. Luth, N.: Automatic identification of music notations. In: Proceedings of the Second International Conference on WEB Delivering of Music (WEDELMUSIC), pp. 203–210 (2002)
20. Marinai, S., Miotti, B., Soda, G.: Bag of characters and som clustering for script recognition and writer identification. Pattern Recognition, International Conference on pp. 2182–2185 (2010)

21. Matsushima, T., S.Ohteru, Hashimoto, S.: An integrated music information processing system: Psb-er. In: In proceedings of 1989 International Computer Music Conference, pp. 191–198. Columbus, Ohio (1989)
22. Mitobe, Y., Miyao, H., Maruyama, M.: A Fast HMM Algorithm Based on Stroke Lengths for On-Line Recognition of Handwritten Music Scores. In: Proceedings of the Ninth International Workshop on Frontiers in Handwriting Recognition, pp. 521–526. IEEE Computer Society (2004)
23. Miyao, H., Maruyama, M.: An online handwritten music symbol recognition system. International Journal on Document Analysis and Recognition **9**(1), 49–58 (2007)
24. Ng, K.: Visual Perception of Music Notation: On-Line and Off-Line Recognition, chap. Optical music analysis for printed music score and handwritten music manuscript, pp. 108–127. Idea Group Inc, Hershey (2004)
25. Perronnin, F.: Universal and adapted vocabularies for generic visual categorization. Pattern Analysis and Machine Intelligence, IEEE Transactions on **30**(7), 1243 – 1256 (2008)
26. Prerau, D.: Computer pattern recognition of standard engraved music notation, phd thesis (1970)
27. Pruslin, D.: Automatic recognition of sheet music, phd thesis (1966)
28. Pugin, L., Burgoyne, J.A., Fujinaga, I.: Goal-directed evaluation for the improvement of optical music recognition on early music prints. In: Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries, pp. 303–304 (2007)
29. Pugin, L., Hockman, J., Burgoyne, J.A., Fujinaga, I.: GAMERA versus ARUSPIX. Two Optical Music Recognition Approaches. In: Proceedings of the 9th International Conference on Music InformationRetrieval, pp. 419–424. Lulu. com (2008)
30. Rebelo, A.: New methodologies towards an automatic optical recognition of handwritten music scores. Master's thesis, Universidade do Porto (Portugal) (2008)
31. Rebelo, A., Capela, G., Cardoso, J.: Optical recognition of music symbols. International Journal on Document Analysis and Recognition **13**, 19–31 (2010)
32. dos Santos Cardoso, J., Capela, A., Rebelo, A., Guedes, C., da Costa, J.: Staff detection with stable paths. IEEE Transactions on Pattern Analysis and Machine Intelligence pp. 1134–1139 (2009)
33. Szwoch, M.: Guido: A musical score recognition system. International Conference on Document Analysis and Recognition **2**, 809–813 (2007)
34. Taubman, G.: Musichand: A handwritten music recognition system. Tech. rep., Brown University (2005)
35. Yoo, J., Kim, G., Lee, G.: Mask Matching for Low Resolution Musical Note Recognition. In: Signal Processing and Information Technology, 2008. ISSPIT 2008. IEEE International Symposium on, pp. 223–226. IEEE (2009)