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

Music staff removal with supervised pixel classification

Jorge Calvo-Zaragoza · Luisa Micó · Jose Oncina

Received: date / Accepted: date

Abstract This work presents a novel approach to tackle the music staff removal. This task is devoted to removing the staff lines from an image of a music score while maintaining the symbol information. It represents a key step in the performance of most Optical Music Recognition systems. In the literature, staff removal is usually solved by means of image processing procedures based on the intrinsics of music scores. However, we propose to model the problem as a supervised learning classification task. Surprisingly, although there is a strong background and a vast amount of research concerning machine learning, the classification approach has remained unexplored for this purpose. In this context, each foreground pixel is labelled as either *staff* or *symbol*. We use pairs of scores with and without staff lines to train classification algorithms. We test our proposal with several well-known classification techniques. Moreover, in our experiments no attempt of tuning the classification algorithms has been made but the parameters were set to the default setting provided by the classification software libraries. The aim of this choice is to show that, even with this straightforward procedure, results are competitive with state-of-the-art algorithms. In addition, we also discuss several advantages of this approach for which conventional methods are not applicable such as its high adaptability to any type of music score.

Keywords Music staff removal · Optical Music Recognition · Pixel classification · Supervised Learning

1 Introduction

Music constitutes one of the main tools for cultural transmission. That is why musical documents have been carefully preserved over the centuries. In an effort to prevent their deterioration, the access to these sources is not always possible. This implies that an important part of this historical heritage remains inaccessible for musicological study. Digitizing this content allows a greater dissemination and integrity of this culture. Furthermore, the massive digitization of music documents opens several opportunities to apply Music Information Retrieval algorithms, which may be of great interest for music analysis. Since the manual transcription of music sources is a long, tedious task —which often requires expert supervision— the development of automatic transcription systems is gaining importance over the last decades [23, 4, 22].

Optical Music Recognition (OMR) can be defined as the ability of a computer to understand the musical information contained in the image of a music score.

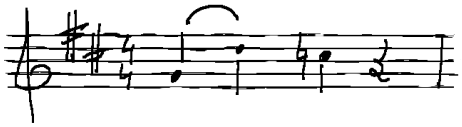
This work has been supported by the Spanish Ministerio de Educación, Cultura y Deporte through a FPU Fellowship (Ref. AP2012–0939) and the Spanish Ministerio de Economía y Competitividad through Project TIMuL (No. TIN2013-48152-C2-1-R, supported by UE FEDER funds).

Jorge Calvo-Zaragoza
Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante, Carretera San Vicente del Raspeig s/n, 03690 Alicante, Spain
Tel.: +349-65-903772
Fax: +349-65-909326
E-mail: jcalvo@dlsi.ua.es

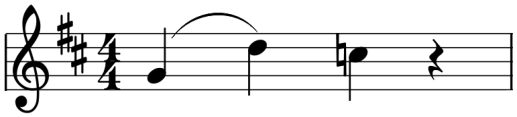
Luisa Micó
Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante, Carretera San Vicente del Raspeig s/n, 03690 Alicante, Spain
E-mail: mico@dlsi.ua.es

Jose Oncina
Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante, Carretera San Vicente del Raspeig s/n, 03690 Alicante, Spain
E-mail: oncina@dlsi.ua.es

The process basically consists in receiving a scanned music score and exporting its musical content to some machine-readable format (Fig. 1). This task can be considered very similar to that of Optical Character Recognition. Nevertheless, its higher complexity and particular notation, in comparison to text, leads to the need of specific developments [1].



(a) Example of input score for an OMR system



(b) Symbolic representation of the input score

Fig. 1 The task of Optical Music Recognition (OMR) is to analyse an image containing a music score to export its musical content to some machine-readable format.

OMR has to deal with many aspects of musical notation, one of which is the presence of the staff, the set of five parallel lines used to define the pitch of each musical symbol. Although these lines are necessary for human readability, they complicate the automatic segmentation and classification of musical symbols. Some works have approached the problem maintaining the staves [24,25,3]; however, a common OMR preprocessing includes the detection and removal of staff lines [28]. This task is aimed at removing the staff lines of the score, maintaining as much as possible the symbol information.

Although staff lines detection and removal may be seen as a simple task, it is often difficult to get accurate results. This is mainly due to problems such as discontinuities, skewing, slant or paper degradation (especially in ancient documents). Given that, the more accurate this process, the better the detection of musical symbols, much research has been devoted to this process, which can be considered nowadays as a research topic by itself.

Notwithstanding all these efforts, the staff-removal stage is still inaccurate and it often produces noise, for example, staff lines not completely removed. Although more aggressive methods that minimize noise can be used, they might produce partial or total loss of some musical symbols. The trade-off between these two aspects, in addition to the accuracy of the techniques, has hitherto led to the inevitable production of errors during this stage. Moreover, the differences among score style, sheet conditions and scanning processes lead researchers to develop some kind of *ad-hoc* method for staff detection and removal, which usually presents little robustness when it is applied to different staves.

From another point of view, the process of removing staff lines can be defined as a classification problem in which, given some foreground pixel, it must be guessed whether that pixel is a part of a *staff* or a *symbol* (*i.e.,* binary classification). Note that addressing the problem in this way, both staff detection and removal can be performed at the same time.

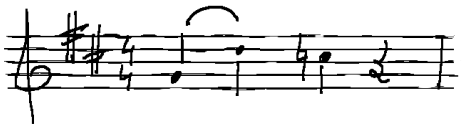
To the best of our knowledge, this approach still remains unexplored. Hence, this work aims at providing a first insight into the staff removal process modelled as a binary classification task. To this end, a set of features based on neighbourhood pixels is extracted at each foreground pixel. At the experimentation stage, several common pattern recognition algorithms will be applied using these features. Our main intention is to show that this simple and general approach deserves further consideration since its performance reaches the level of state-of-art methods while offering several advantages that the others can not.

This paper is organized as follows: Section 2 presents background on staff detection and removal; Section 3 describes our approach to model the process as a classification task; Section 4 contains the experimentation performed and the results obtained; Section 5 discusses the pros and cons of our approach, and some additional considerations; and finally, Section 6 concludes the current work.

2 Background

Due to the complexity of music notation, OMR systems rely on music staff removal algorithms to perform the most relevant task of symbol isolation and segmentation [26]. Note that this process should not only detect staff lines but also remove them in such a way that musical symbols remain intact (see Fig. 2).

Unfortunately, this removal stage is hardly ever perfect. The need of eliminating every part of the staff often leads to delete some parts of the musical symbols, which produces unavoidable errors in posterior stages. The trade-off between keeping symbols and removing staff lines leads to inevitable production of extraction and classification errors later. That is why several methods have been proposed to tackle this process. A good comparative study, including a taxonomy of the different approaches, can be found in the work of Dalitz et al. [7].



(a) Example of input score for an OMR system

(b) Input score after staff removal

Fig. 2 Example of an accurate staff removal process.

In the last years, however, new strategies have been developed: Cardoso et al. [30] proposed a method that considers the staff lines as connecting paths between the two margins of the score. Then, the score is modelled as a graph so that staff detection is solved as a maximization problem. This strategy was improved and extended to be used on grey-scale scores [27]; Dutta et al. [10] developed a method that considers the staff line segment as a horizontal connection of vertical black runs with uniform height, which are validated using neighbouring properties; in the work of Piatkowska et al. [21], a Swarm Intelligence algorithm was applied to detect the staff line patterns; Su et al. [31] start estimating properties of the staves like height and space, then, they tried to predict the direction of the lines and fitted an approximate staff, which was posteriorly adjusted; Geraud [13] developed a method that entails a series of morphological operators directly applied to the image of the score to remove staff lines; and Montagner et al. [19] proposed to learn image operators, following the work of Hirata [17], whose combination was able to remove staff lines. Others works have addressed the whole OMR problem by developing their own, case-directed staff removal process [29,32].

The current performance of staff removal methods can be checked in the *GREC/ICDAR 2013 Staff Removal Competition* [34,12]. This competition makes use of the CVC-MUSCIMA database [11], which contains handwritten music score images with a perfect ground-truth on staff removal. Many of the most advanced methods showed a decreasing accuracy when different distortions were applied to the input scores. Indeed, the same behaviour may be expected by methods especially suitable for some type of score that are subsequently applied to very different conditions. Taking into account the vast variety of music manuscripts —which is even wider considering old music— there is a need of developing staff removal methods that are able to deal with any kind of score.

In our work, we propose to model the staff removal stage as a pixel classification problem. That is, extract features from each foreground pixel and take a decision about keeping or removing it based on supervised learning classification techniques. Therefore, the accuracy of the method lies in data instead of in selecting the appropriate series of image processing steps. Although it may be worse in the cases in which specific staff removal algorithms have been developed, it allows us to present a robust approach since it can be effective in any type of score as long as labelled data is available. The strategy proposed is described in next section.

3 A Classification Approach for Staff Lines Removal

As depicted above, several procedures for the staff detection and removal stage have been proposed. Although most of them are able to achieve a very good performance in many cases, they are far from optimal when the style of the score is changed. The intention of our strategy is to present a new method that is able to adapt to the actual score style as long as learning data is available.

To handle this issue, we propose to follow a supervised learning approach. That is, the task is based on building a classification model using a training sample with labelled data. After that, the model is able to receive new unseen samples and determine the class label [9].

In our context, given an image depicting a score, we extract a labelled set of features from each foreground pixel. These features are used to train a classification algorithm. At test phase, each of these pixels is classified between *symbol* or *staff*. Then, depending on what it is pursued —either staff detection or staff removal— it is removed from the image those pixels classified as symbol or those classified as staff. Without loss of generality, we shall assume from now on that our objective is the staff removal stage since it is the common preprocessing required in OMR systems.

In this work the features of each pixel of interest consist of the values of its neighbouring region. We believe that the surroundings of each pixel contains contextual information that can be discriminative enough for this task. Furthermore, this contextual information can help to avoid misclassification due to noise or small deformations of the image.

We shall assume that the input score has been binarized previously, as it is usual in this field. Nevertheless,

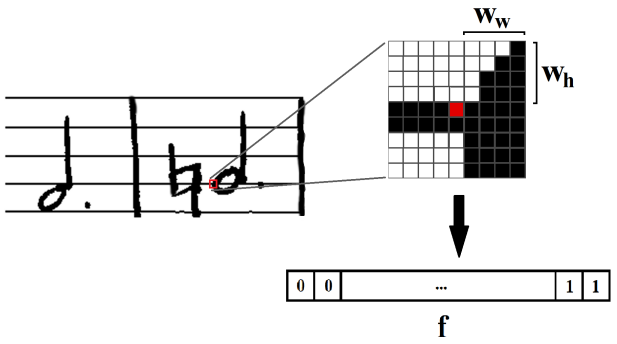


Fig. 3 Example of feature extraction considering $w_w = w_h = 4$. Cell in red represents the pixel from which features are being extracted.

our feature extraction is not restricted to binary images but it could be applied to any type of image.

Formally speaking, let $I : \mathbb{N} \times \mathbb{N} \rightarrow \{0, 1\}$ define the input score image. We use w_h and w_w to denote two integer values defining the opening of the neighbouring region in each dimension (horizontal and vertical, respectively). Given a position (i, j) of the input image, a set of $(2\,w_h+1)(2\,w_w+1)$ features ($\mathbf{f}_{i,j}$) is considered taking the values of the neighbourhood region centred at (i, j). That is, $\mathbf{f}_{i,j} = \{I(x, y) : |i - x| \leq w_h \wedge |j - y| \leq w_w\}$. Then, the values contained within this set are concatenated following some specific order (*e.g.,* by columns) to obtain a proper feature vector. This process is illustrated in Fig. 3.

To obtain the training set, the feature extraction process is applied to the foreground pixels of a labelled dataset of scores with and without staff lines. Given a pixel in the position (i, j), the feature extraction is applied in the score that contains staff lines (*i.e.,* in the original one). After that, the value in the position (i, j) of the score without staff lines is used to obtain the actual label between *staff* or *non-staff*.

This training set is used to feed a supervised learning classifier. Then, when an input score is received, this classifier will be able to take a decision about each of its foreground pixels. If it is classified as *staff*, the pixel will be removed from the image.

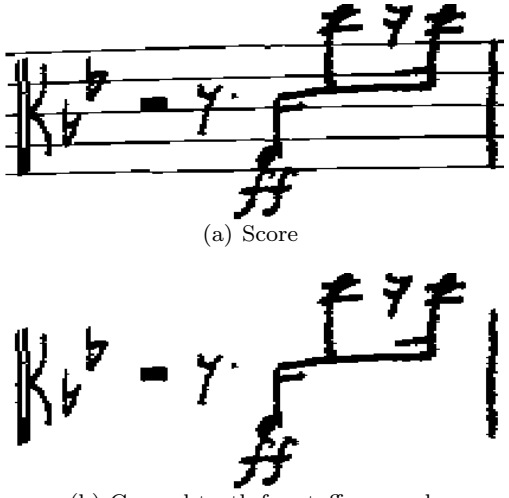
It should be emphasised that the intention of this work is not to find the most suitable pair of features and classification algorithm, but to show that this approach allows dealing with the staff removal stage even with a very straightforward classification step. Pursuing the best configuration could cause results depend more on these aspects than on the approach itself. Thus, a more comprehensive feature extraction and classification research fall outside the actual intention of this work.

Next section will present experimentation with some common classification algorithms considering several values of w_h and w_w.

4 Experimentation

This section details the experimentation carried out to test our proposal. Taking advantage of the *GREC/ICDAR 2013 Staff Removal Competition*, whose dataset was publicly available[1], we are going to follow the same experimental set-up to assure a fair comparison with state-of-art developments.

The data used in this contest is organized in train and test sets, with 4 000 and 2 000 samples respectively. The test set is further divided into three subsets (TS1, TS2, and TS3) based on the deformations applied to the scores. Each sample consists of an image of a handwritten score in both binary and grey-scale with its corresponding ground-truth (the score without staves). We shall use here the binary ones. Figure 4 shows a piece from a score of that set. The number of foregrounds pixels per score is around 500 000 with 200 000 staff pixels, both on average.



(a) Score

(b) Ground-truth for staff removal

Fig. 4 Piece of sample from the *GREC/ICDAR 2013 Staff Removal Competition* dataset.

The training set will be used to learn to distinguish between staff and symbol pixels by the classification algorithms. Due to the large amount of data available, it is infeasible to handle it completely. Thus, at each instance only one score of the training set, chosen randomly, will be used. If we also consider that one score

[1] `http://dag.cvc.uab.es/muscima/`

contains around 500 000 foreground pixels, this is still too much information to use as training set.

We must bear in mind that the whole set of foreground pixels may be used to train the classifiers. Then, to further reduce the size of the training sample, the Condensing algorithm [16] was applied. This algorithm removes the samples that are not considered relevant enough for the classification task. After that, the average size of the training sample was around 20 000. In other words, only 4 % of the foreground pixels of one score randomly selected have been used as training, which constitutes 0.001 % of the available training information.

On the other hand, test set will be used to assess the results achieved. As in the competition, the performance metric will be the F_1 score or *F-measure*:

$$F_1 = \frac{2\,TP}{2\,TP + FP + FN}$$

where TP, FP and FN stand for true positives (staff pixels classified as staff), false positives (symbol pixels classified as staff) and false negatives (staff pixels classified as symbol), respectively.

4.1 Classification techniques

For the classification task, many supervised learning algorithm can be applied. In this work, we are going to consider the following methods:

- Nearest Neighbour (NN) [6]: given a distance function between samples, this algorithm proposes a label of the input by querying its nearest neighbour of the training set. The Euclidean distance was used for our task.
- Support Vector Machine (SVM) [33]: it learns a hyperplane that maximises the distance to the nearest samples (support vectors) of each class. It makes use of Kernel functions to handle non-linear decision boundaries. In our case, a *Radial Basis Function* kernel was chosen.
- Random Forest (RaF) [2]: it builds an ensemble classifier by generating several random decision trees at the training stage. The final output is taken by combining the individual decisions of each tree.

The methods described above have been applied using the Waikato Environment for Knowledge Analysis (WEKA) library [15], each one with their default parametrization unless where it has been told otherwise. Since our interest is not focused on finding the best classifier for this task, but to emphasise the supervised learning approach, we did not pursue the optimal tuning of these classifiers.

Table 1 Average F_1 score (%) achieved by the different classification techniques in combination with different values of neighbouring squared region over the three test subsets. Bold values indicate the best results, on average, at each subset. The average results obtained with each set of features are also showed.

Test set	Classifier	Features			
		9	25	49	81
TS1	NN	68.34	86.10	89.69	91.07
	SVM	40.72	87.14	93.95	**94.10**
	RaF	68.06	90.12	93.52	93.89
TS2	NN	77.32	90.05	95.24	96.06
	SVM	51.22	97.02	**98.11**	98.08
	RaF	76.46	93.86	96.95	97.78
TS3	NN	71.56	86.23	89.33	90.58
	SVM	48.07	87.81	93.92	**94.00**
	RaF	71.15	90.55	93.23	93.39
Average		63.43	89.87	93.77	94.32

4.2 Results

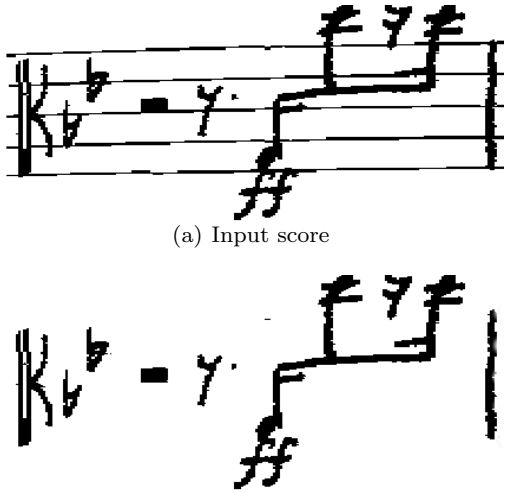
For our experiments we shall restrict ourselves to consider squared neighbouring regions. Concretely, regions with windows of length 1, 2, 3 and 4 in each direction centred at the pixel. Therefore, each pixel of interest is defined by 9, 25, 49 and 81 binary features, respectively. Results achieved are shown in Table 1.

An initial remark is that the number of features has a stronger influence in the results than the algorithm used. For instance, classifiers showed a poor performance when 9 features are considered but they increase noticeably the accuracy with 25 features. It is important to stress that each configuration outperforms results of any other configuration with less features, with independence of the algorithm used. This is also reported in the last row of the table, in which the average improvement obtained by increasing the number of features is depicted.

Results seem to be stabilized within the two highest number of features considered. Thereby including more than 81 is not expected to improve accuracy significantly. In addition, increasing the number of features may imply some drawbacks such as efficiency in both learning and testing phase.

Regarding classification techniques, SVM achieves best results for both each feature extraction considered and each test subset, although its difference with RaF is hardly significant. In turn, NN does present a lower accuracy than the others. Specifically, SVM with 81 is reported as the best configuration, on average, and it is also the best result in two of the three corpora used.

Figure 5 shows an example of the behaviour provided by our algorithm (SVM with 81 features), in which a great accuracy is achieved. Surprisingly, misclassification is not found in the edges between symbol



(a) Input score

(b) Score after staff removal

Fig. 5 Example of a staff removal process using SVM classifier, 81 features per pixel and only one *condensed* score as training set.

and staff, but it mainly occurs along the staff lines. In fact, looking in more detail, very few symbol pixels are removed. It should be stressed that most of these remaining mistakes could be hopefully corrected by means of a post-processing step.

To analyse our performance against state-of-art methods, Table 2 shows a summary of the results achieved in the staff removal competition for each test set. These sets comprise different deformations applied over original scores: 3D distortions in TS1, local noise in TS2, and both 3D distortion and local noise in TS3. For a detailed description about each participant and the deformation models applied, reader is referred to the report of the competition [34]. Our best average configuration (SVM with 81 features) is also included for comparison.

Most of the methods proposed in the contest follow a two step approach: first, an estimation of the position of the staff lines; then, staff lines removal while keeping symbol information. This second step is what usually produces the accuracy loss, since it is difficult to distinguish symbol pixels over a staff line. On the contrary, our method is directly focused on the final task without a first estimation of staff lines.

According to the results, our method shows the best accuracy against local noise (TS2). This is probably because local noise is less harmful for our feature extraction and classification. In turn, they are less generalizable to deal with 3D distortions (TS1 and TS3), for which our approach suffers some accuracy loss. Although we only achieve the highest score in one of the two subsets considered, our results are quite competitive as differences among best results and those obtained by our method are very small. In addition, our method surpasses many of the participants in the contest in all sets considered. It should be noted that not only this configuration is competitive but also most of the configurations showed in Table 1, even with a little set of features. Moreover, we must also remember that for obtaining these results only 0.001 % of all available training information was used.

Table 2 F_1 comparison between the best tuning of our method and the participants in the staff removal contest. Best values, on average, achieved on each subset are highlighted.

Method	TS1	TS2	TS3
TAU	85.72	81.72	82.29
NUS	69.85	96.25	67.43
NUASI-lin	94.77	94.76	93.81
NUASI-skel	94.11	93.67	92.78
LRDE	**97.73**	96.86	**96.98**
INESC	89.29	97.72	88.52
Baseline	87.01	96.91	89.90
Our method	94.10	**98.08**	94.00

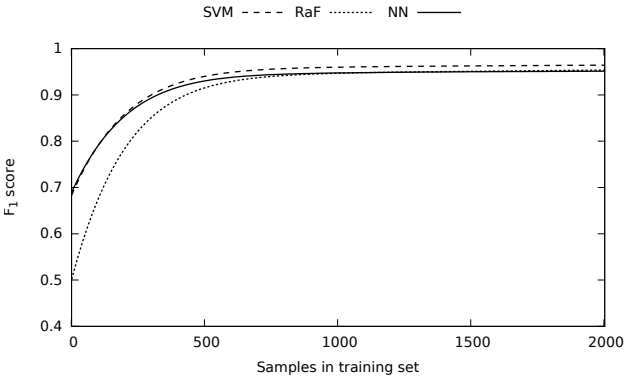


Fig. 6 Performance of the classifiers using 81 features with respect to the amount of training samples.

Finally, we focused on assessing whether the amount of data used to train the classifiers has a strong impact on the results. To this end, another experiment has been performed in which the number of training samples is iteratively increased, using a random subset of the training scores as validation set. As mentioned above, the specific size of the training set in our previous experiments is given by Condensing algorithm, which keeps around 20 000 samples, on average. Figure 6 shows the curves of such experiment extracting 81 features per pixel (the highest value considered in our experiments). It can be seen that the performance is already stable when classifiers are trained with 2000 samples. This leads to the insight that results are not expected to improve significantly if more data were considered.

Given all of above, we consider that our proposal should merit high interest since its performance is com-

petitive using a simple strategy that has not been studied so far. Next section extends the implications that our method has in the ongoing research on staff removal, supported by the results obtained.

5 Discussion

Since the work presented here is the first approach to the staff removal task as a pixel classification problem, it opens several lines of discussion that should be addressed.

The first thing to remark is that the performance of our method is very competitive, although it does not significantly outperform all the already proposed ones. While this fact may question the usefulness of the proposal, relevant additional advantages are shown. First of all, it is simple, easy-to-use and does not require additional knowledge of the field. In addition, a fine tuning of the classifiers parameters, as well as using some kind of advanced feature extraction, clearly represent room for accuracy improvement.

Unfortunately, this method has also drawbacks that deserve consideration in future developments. For instance, approaching the task from a classification point of view is very expensive. Regardless the specific classifier speed, each foreground pixel of the score entails a classification process. Therefore, our method will be usually slower than conventional image processing methods.

From the learning-driven process point of view, the staff removal stage is as robust as its training set. That is, the process can be accurate if we have enough data of the target type of score. Foreground information such as hand-written notation or noise can also be addressed simultaneously as long as they appear in the training data. Furthermore, this paradigm allows the method to be adapted to any type of music score, even those quite different such as Gregorian chant or guitar tablatures. In those cases, classical methods may fail because of the high variation with respect to classical notation or the variable number of staff lines.

To serve as an example, we have carried out a simple *proof of concept* experiment to compare the adaptiveness of our proposal against a classical one. The experiment is focused on early music manuscripts so as to analyse the behaviour of the methods when dealing with quite different musical scores.

In order to feed the supervised learning classifier, we have manually labelled a single line of staff of this type (see Fig. 7). Note that we are just using a very small piece as training set, which is expected to be available with small effort.

As a representative of classical image processing strategies we have chosen the LRDE method, since it depicted the best performance in the contest. Its publicly available online demo[2] has been used for this test.

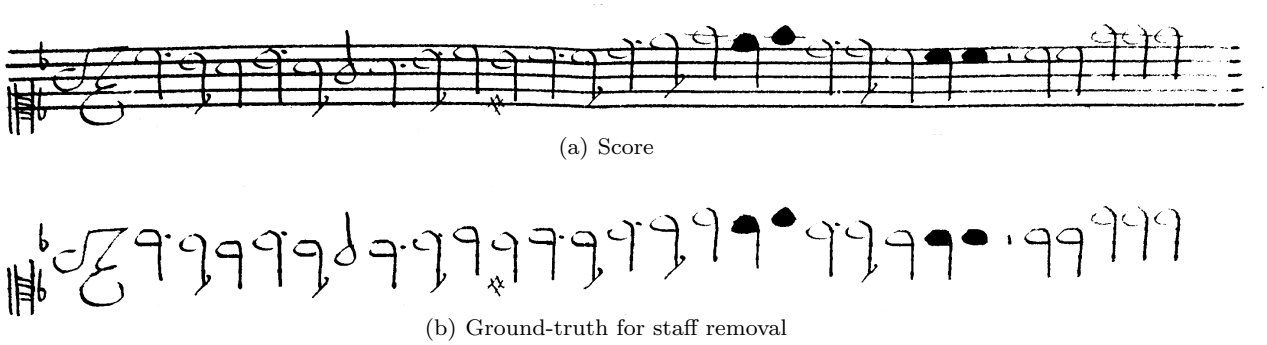
Figure 8 shows the results of a staff removal process applying both our proposal and this method to an early music piece of score. For the sake of further analysis, our method is trained with both specific data and data of CVC-MUSCIMA. It is important to remark that the LRDE method is not able to remove accurately staff lines in spite of being one of the best in the contest. On the other hand, our method achieves a very poor performance if data is not appropriate, as depicted in Fig. 8(c). However, if specific data is used, results are fairly accurate (Fig. 8(d)). Although this comparison may not be totally fair, it clearly illustrates some drawback of developing image procedures to remove staff lines in contrast to a learning-based approach.

5.1 Further considerations

In addition to the advantages discussed previously, considering staff removal as a supervised classification problem allows us to explore many other paradigms that could be profitable for this task:

- Online Learning: new data may be available through the use of the system [8]. For instance, when user corrects OMR mistakes, the information could be analysed to extract new labelled pixels for the staff removal process. This case may be useful when it is assumed that the data of the image in process is more relevant than the training set itself.
- Active Learning: if it is assumed that a user must be supervising the OMR task, the system could query about the category (staff or symbol) of some piece of the score. The main goal is to reduce the user effort in the whole process, and therefore some queries may be needed to avoid many of the potential mistakes in the classification stage [14].
- One-class classification: since the staff removal may entail an imbalanced binary classification with respect to the training data available, it could also be modelled as a one-class classification problem [18]. This case seems to be very interesting because it would need less data to train (just one of the two classes considered, the one whose data is more available) and some strategies could be applied to automatically extract labelled data of that class from score images.
- Deep Learning: taking into account the huge amount of labelled data present in this task, this paradigm

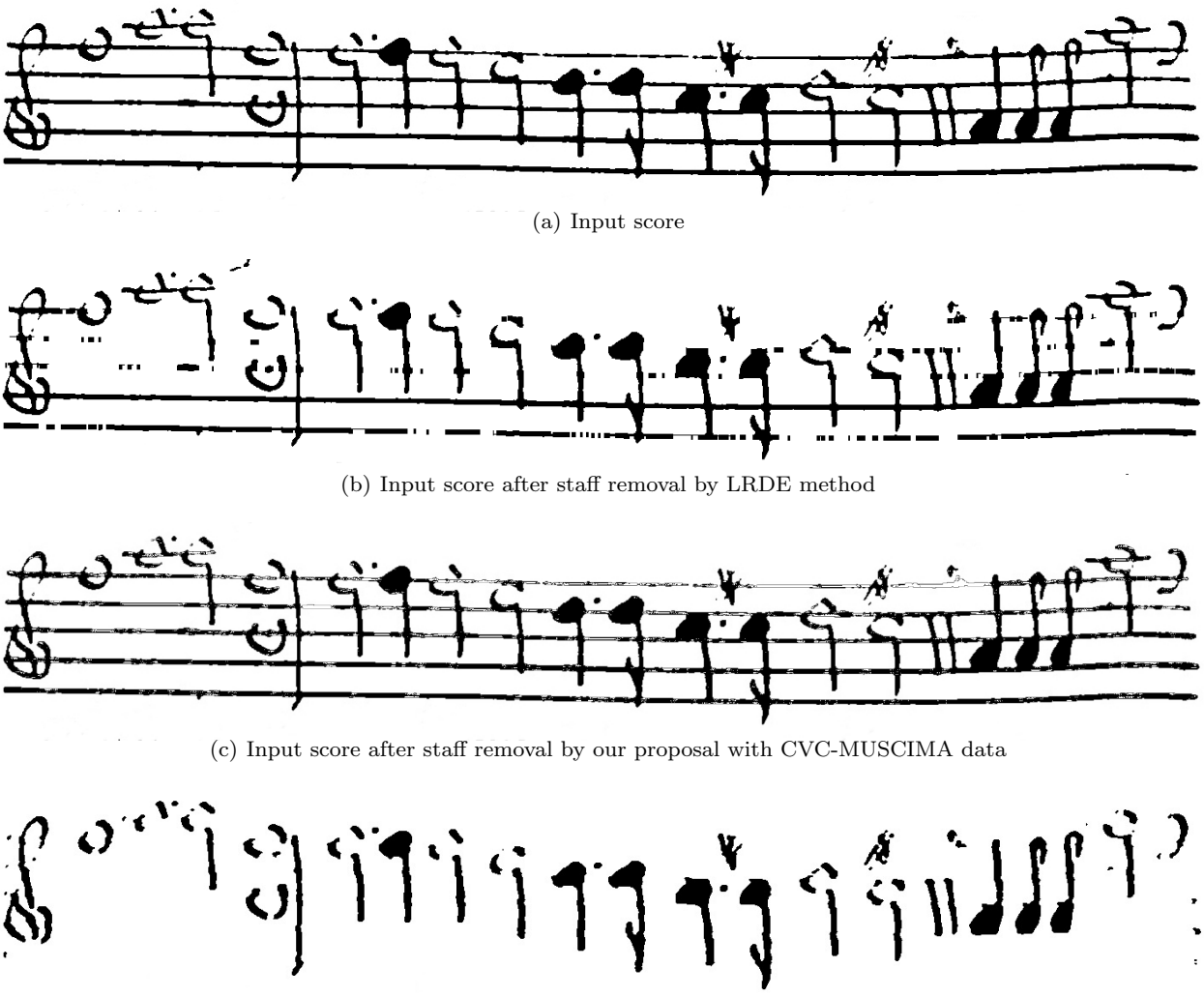
[2] `https://olena.lrde.epita.fr/demos/staff_removal.php`



(a) Score

(b) Ground-truth for staff removal

Fig. 7 Training set used for the *proof of concept* experiment over early music scores.



(a) Input score

(b) Input score after staff removal by LRDE method

(c) Input score after staff removal by our proposal with CVC-MUSCIMA data

(d) Input score after staff removal by our proposal with specific data

Fig. 8 Performance of LRDE method and our proposal (SVM classifier and 81 features per pixel) with general and specific data over an ancient score of early music.

may learn the high-level representation inherent to each piece of the score to learn to distinguish between symbol and staff pixels more accurately. Convolutional Neural Networks have been reported to be especially suitable for performing such a task [5].

These points represent ideas that could be implemented to improve the process so that it becomes more adaptive, efficient and/or effective. Nevertheless, it should be noted that most of these paradigms can not be applied if conventional methods for staff removal are used.

On the other hand, one of the main obstacles in the preprocessing of degraded documents is the binarization step. However, the method proposed in this work could be trained to deal with grey-level images, although it would represent a different task. Since background pixels would have to be classified as well, the complexity of the process would be increased drastically.

For all the reasons above, we believe that this approach is worthwhile in its current form since the performance achieved is comparable to state-of-art with a very straightforward procedure. Moreover, it is specially interesting when considering all the research avenues and opportunities opened, some of which could lead to a significantly higher performance than that obtained by the methods proposed so far.

6 Conclusions

In this work we presented a novel approach for the staff-removal stage, a key pre-processing step in the performance of most OMR systems. Our strategy models the task as a supervised learning classification problem, in which each foreground pixel is classified as *staff* or *symbol* using raw neighbouring pixels as features.

In our experiments, the feature set was demonstrated to be more relevant than the specific classifier in the accuracy results. SVM classifier, considering 81 features, reported the best results on average. In comparison with other state-of-art staff removal processes, our strategy showed a very competitive performance, even achieving the best results in some cases, using a very small piece of the training information. A proof of concept experiment over early music scores has also been carried out as an example of the robustness of our method. Therefore, this novel approach deserves further consideration in the field since it also opens several opportunities for which conventional methods are not applicable.

As future work some effort should be devoted to overcoming the problem of getting enough data to train the classifiers. For instance, the conditions of the actual sheet —such as scale, deformation and so on— could be learned online. Then, the same conditions could be applied to a reference dataset so that specific labelled data is obtained for each type of score. The use of adaptive techniques for domain adaptation or transfer learning is another way to deal with this issue [20]. Similarly, considering an interactive OMR system, staff-removal learning could be improved through user interaction.

Moreover, there is still plenty of room for improvement regarding the classification process such as seeking a better feature set or using other advanced techniques for supervised learning. Speeding-up the process may be also of great interest. For instance, by classifying a relatively small block of the score at a once, instead of querying every single pixel of the image.

References

1. Bainbridge, D., Bell, T.: The challenge of optical music recognition. Computers and the Humanities **35**(2), 95–121 (2001). DOI 10.1023/A:1002485918032
2. Breiman, L.: Random forests. Machine Learning **45**(1), 5–32 (2001). DOI 10.1023/A:1010933404324
3. Calvo-Zaragoza, J., Barbancho, I., Tardón, L.J., Barbancho, A.M.: Avoiding staff removal stage in optical music recognition: application to scores written in white mensural notation. Pattern Anal. Appl. **18**(4), 933–943 (2015)
4. Carter, N.P.: Segmentation and preliminary recognition of madrigals notated in white mensural notation. Mach. Vis. Appl. **5**(3), 223–229 (1992). DOI 10.1007/BF02627000
5. Ciresan, D., Meier, U., Schmidhuber, J.: Multi-column deep neural networks for image classification. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3642–3649. IEEE (2012)
6. Cover, T., Hart, P.: Nearest neighbor pattern classification. IEEE Transactions on Information Theory **13**(1), 21–27 (1967). DOI 10.1109/TIT.1967.1053964
7. Dalitz, C., Droettboom, M., Pranzas, B., Fujinaga, I.: A comparative study of staff removal algorithms. IEEE Trans. Pattern Anal. Mach. Intell. **30**(5), 753–766 (2008). DOI 10.1109/TPAMI.2007.70749
8. Diethe, T., Girolami, M.: Online learning with multiple kernels: A review. Neural Comput. **25**(3), 567–625 (2013)
9. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2 edn. John Wiley & Sons, New York, NY (2001)
10. Dutta, A., Pal, U., Fornes, A., Llados, J.: An efficient staff removal approach from printed musical documents. In: 2010 20th International Conference on Pattern Recognition (ICPR), pp. 1965–1968 (2010)
11. Fornés, A., Dutta, A., Gordo, A., Lladós, J.: CVC-MUSCIMA: a ground truth of handwritten music score images for writer identification and staff removal. International Journal on Document Analysis and Recognition (IJDAR) **15**(3), 243–251 (2012)
12. Fornés, A., Kieu, V.C., Visani, M., Journet, N., Dutta, A.: The ICDAR/GREC 2013 music scores competition: Staff removal. In: 10th International Workshop on Graphics Recognition, Current Trends and Challenges GREC 2013, Bethlehem, PA, USA, August 20-21, 2013, Revised Selected Papers, pp. 207–220 (2013)

13. Géraud, T.: A morphological method for music score staff removal. In: Proceedings of the 21st International Conference on Image Processing (ICIP), pp. 2599–2603. Paris, France (2014)
14. Gosselin, P., Cord, M.: Active learning methods for interactive image retrieval. IEEE Transactions on Image Processing **17**(7), 1200–1211 (2008)
15. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: An update. SIGKDD Explor. Newsl. **11**(1), 10–18 (2009). DOI 10.1145/1656274.1656278
16. Hart, P.: The condensed nearest neighbor rule (corresp.). IEEE Transactions on Information Theory **14**(3), 515–516 (1968). DOI 10.1109/TIT.1968.1054155
17. Hirata, N.S.T.: Multilevel training of binary morphological operators. IEEE Trans. Pattern Anal. Mach. Intell. **31**(4), 707–720 (2009)
18. Khan, S.S., Madden, M.G.: One-class classification: taxonomy of study and review of techniques. The Knowledge Engineering Review **29**, 345–374 (2014)
19. Montagner, I.d.S., Hirata, R., Hirata, N.S.: A machine learning based method for staff removal. In: Pattern Recognition (ICPR), 2014 22nd International Conference on, pp. 3162–3167 (2014). DOI 10.1109/ICPR.2014.545
20. Patel, V.M., Gopalan, R., Li, R., Chellappa, R.: Visual domain adaptation: A survey of recent advances. IEEE Signal Process. Mag. **32**(3), 53–69 (2015)
21. Piatkowska, W., Nowak, L., Pawlowski, M., Ogorzalek, M.: Stafflines pattern detection using the swarm intelligence algorithm. In: L. Bolc, R. Tadeusiewicz, L.J. Chmielewski, K. Wojciechowski (eds.) Computer Vision and Graphics, *Lecture Notes in Computer Science*, vol. 7594, pp. 557–564. Springer Berlin Heidelberg (2012)
22. Pinto, J.R.C., Vieira, P., Ramalho, M., Mengucci, M., Pina, P., Muge, F.: Ancient music recovery for digital libraries. In: Proceedings of the 4th European Conference on Research and Advanced Technology for Digital Libraries, ECDL '00, pp. 24–34. Springer-Verlag, London, UK, UK (2000)
23. Pruslin, D.: Automatic recognition of sheet music. Sc.d. dissertation, Massachusetts Institute of Technology (1966)
24. Pugin, L.: Optical music recognition of early typographic prints using hidden markov models. In: ISMIR 2006, 7th International Conference on Music Information Retrieval, Victoria, Canada, 8-12 October 2006, Proceedings, pp. 53–56 (2006)
25. Ramirez, C., Ohya, J.: Automatic recognition of square notation symbols in western plainchant manuscripts. Journal of New Music Research **43**(4), 390–399 (2014)
26. Rebelo, A., Capela, G., Cardoso, J.S.: Optical recognition of music symbols. International Journal on Document Analysis and Recognition (IJDAR) **13**(1), 19–31 (2010)
27. Rebelo, A., Cardoso, J.: Staff line detection and removal in the grayscale domain. In: 2013 12th International Conference on Document Analysis and Recognition (ICDAR), pp. 57–61 (2013). DOI 10.1109/ICDAR.2013.20
28. Rebelo, A., Fujinaga, I., Paszkiewicz, F., Marçal, A.R.S., Guedes, C., Cardoso, J.S.: Optical music recognition: state-of-the-art and open issues. IJMIR **1**(3), 173–190 (2012). DOI 10.1007/s13735-012-0004-6
29. Rossant, F., Bloch, I.: Robust and adaptive omr system including fuzzy modeling, fusion of musical rules, and possible error detection. EURASIP Journal on Applied Signal Processing **2007**(1), 160–160 (2007)
30. dos Santos Cardoso, J., Capela, A., Rebelo, A., Guedes, C., Pinto da Costa, J.: Staff detection with stable paths. IEEE Trans. Pattern Anal. Mach. Intell. **31**(6), 1134–1139 (2009)
31. Su, B., Lu, S., Pal, U., Tan, C.: An effective staff detection and removal technique for musical documents. In: 2012 10th IAPR International Workshop on Document Analysis Systems (DAS), pp. 160–164 (2012). DOI 10.1109/DAS.2012.16
32. Tardón, L.J., Sammartino, S., Barbancho, I., Gómez, V., Oliver, A.: Optical music recognition for scores written in white mensural notation. EURASIP J. Image and Video Processing **2009** (2009)
33. Vapnik, V.N.: Statistical learning theory, 1 edn. Wiley (1998)
34. Visani, M., Kieu, V., Fornes, A., Journet, N.: ICDAR 2013 Music Scores Competition: Staff Removal. In: 2013 12th International Conference on Document Analysis and Recognition (ICDAR), pp. 1407–1411 (2013). DOI 10.1109/ICDAR.2013.284