A MORPHOLOGICAL METHOD FOR MUSIC SCORE STAFF REMOVAL

Thierry Géraud

EPITA Research and Development Laboratory (LRDE)
14–16, rue Voltaire, FR-94276 Le Kremlin-Bicêtre, France
thierry.geraud@lrde.epita.fr

ABSTRACT

Removing the staff in music score images is a key to improve the recognition of music symbols and, with ancient and degraded handwritten music scores, it is not a straightforward task. In this paper we present the method that has won in 2013 the staff removal competition, organized at the International Conference on Document Analysis and Recognition (ICDAR). The main characteristics of this method is that it essentially relies on mathematical morphology filtering. So it is simple, fast, and its full source code is provided to favor reproducible research.

***Index Terms*—** Document Image Analysis; Music Score; Mathematical Morphology; Filtering.

1. INTRODUCTION

The quality in recognition of music symbols is often subordinated to a preliminary step consisting in detecting and removing staff lines. With ancient and degraded handwritten music scores, removing properly the staff lines is challenging. A competition organized at the International Conference on Document Analysis and Recognition (ICDAR) in 2013[1] has intended to compare some state-of-the-art methods. To that aim, a recent database of score images has been used, CVC-MUSCIMA [1], featuring the remarkable properties that the images are handwritten, that they are degraded with different methods and various intensity, and that a complete ground-truth dataset exists. There were 9 methods in competition (see [2] and [3] for details). The method presented in this paper won the competition, with an average F-measure of **0.97** (precision 0.98 and recall 0.96) and an average accuracy of **0.998**. Its main characteristics is that it essentially relies on mathematical morphology filtering.

As forewords, we have to admit that the work presented in this paper is not "innovative", in the sense that it relies on very simple morphological operators. The only valuable contribution of this paper is that it provides the community with a new effective method (processing chain) to remove staff lines, along with the availability of its source code. Nevertheless this method has two main advantages: it is robust to image degradations (the competition tried to focus on that particular evaluation), and it is very simple.

Mathematical morphology has been defined to deal with shapes in images. At its origin, a collection of operators and their related properties have been bring to the fore to process binary images, i.e., sets. Considering a space E (typically a subset of $\mathbb{Z}^2$), let us denote by S_e the translation of the set $S \subset E$ by an element $e \in E$. Given a structuring element B (a tiny set, usually centered), the *dilation* of a set X by B is the Minkowski sum of those sets:

$$\delta_B(X) \;=\; X \oplus B \;=\; \bigcup_{b \in B} X_b$$

and the *erosion* of X by B is:

$$\varepsilon_B(X) \;=\; \{\, p \mid B_p \subseteq X \,\} \;=\; E \setminus \delta_B(E \setminus X).$$

Those two dual operators lead to a bestiary of morphological operators working on sets. Let us just recall a few of them:

closing	$\varphi_B \,=\, \varepsilon_B \circ \delta_B$
opening	$\gamma_B \,=\, \delta_B \circ \varepsilon_B$
hit-or-miss	$\upsilon_{(B_1,B_2)}(X) \,=\, \varepsilon_{B_1}(X) \cap \varepsilon_{B_2}(E \setminus X)$
thickening	$\upsilon^+_{(B_1,B_2)}(X) \,=\, X \cup \upsilon_{(B_1,B_2)}(X)$
thinning	$\upsilon^-_{(B_1,B_2)}(X) \,=\, X \setminus \upsilon_{(B_1,B_2)}(X)$
gradients	$\nabla_B(X) \,=\, \delta_B(X) \setminus \varepsilon_B(X)$
	$\nabla^+_B(X) \,=\, \delta_B(X) \setminus X$ (external grad.)
	$\nabla^-_B(X) \,=\, X \setminus \varepsilon_B(X)$ (internal grad.)
top-hats	$\tau^+_B(X) \,=\, \varphi_B(X) \setminus X$ (black t.-h.)
	$\tau^-_B(X) \,=\, X \setminus \gamma_B(X)$ (white t.-h.)
	$\tau_B(X) \,=\, \varphi_B(X) \setminus \gamma_B(X)$

Just note that, when the structuring element B is a segment (horizontal or vertical) or a rectangle, those operators have very efficient implementations. All the operators given above are actually very basic, easy to understand, and their use is effective in practical cases.

From a mathematical morphology practitioner point of view, knowing the many tools of that field, it is rather obvious that mathematical morphology is very well-suited to perform staff removal. Furthermore, from this same point of view, the processing chain presented in this paper is rather straightforward and simple.

[1] http://www.icdar2013.org/program/competitions

(a) Input image. (b) Step 1: permissive hit-or-miss.

(c) Step 2: horizontal median filter. (d) Step 3: horizontal reconstruction.

(e) Step 4: cleaning. (f) Step 5: after line selection (contour superimposed).

(g) Step 6: output, after a local vertical median filter. (h) Ground truth.

Fig. 1: Step-by-step staff removal: the same image excerpt is depicted (on the left of each sub-figure) with a zoom on a bottom-left located part (on the right of each sub-figure).

This paper is organized as follows. Section 2 describes the proposed staff removal method. Section 3 compares it with some other state-of-the-art approaches. The short Section 4 is dedicated to the important matter of reproducible research. Last we conclude in Section 5.

2. DESCRIPTION

Our method to remove the staff is a linear processing chain composed of six steps. They are presented here by the following subsections and are illustrated in Figure 1. Shortly put the rationale behind the processing chain is to extract a mask (steps 1 to 4) that fits well the staff lines, to filter out the components of this mask that are due to the presence of ties (step 5), and eventually to remove the staff lines without affecting the music symbols (step 6):

1. extract chunks of staff lines;
2. regularize their shapes;
3. extend the chunks horizontally;
4. correct some defects;
5. select staff lines, i.e., get rid of tie lines;
6. reconstruct an image without staff lines.

Some details are deliberately omitted in the description of those processing steps below. That is specifically the case of the parameters that we use; they were set with respect to the resolution of the input score images—just note that the method is rather robust to some reasonable variations of those parameters. All those (important) details can be accessible since we provide the reader with the full implementation of our method (see Section 4).

2.1. A Permissive Hit-or-Miss to Obtain Chunks

The intent of this first step is to get something to work with, that is as close to the final expected result as we can imagine: we want to extract chunks of staff lines. For that, we rely on a modified hit-or-miss, defined with B_1 being a horizontal segment centered at $\mathbf{0}$, and with B_2 being two horizontal segments respectively shifted above and below $\mathbf{0}$ (so $B_1 \cup B_2$ looks like the symbol $\equiv$). Applying the hit-or-miss $\upsilon_{(B_1,B_2)}(X) \;=\; \varepsilon_{B_1}(X) \,\cap\, \varepsilon_{B_2}(E \setminus X)$ means that we retain only pixels surrounded horizontally by the object (thanks to $\varepsilon_{B_1}(X)$) and where the horizontal surroundings below and above fall in the background (thanks to $\varepsilon_{B_2}(E \setminus X)$); we thus keep only pixels belonging to thin horizontal lines. To be robust at the same time to noise, to staff curvature, and to the presence of other objects (notes, barlines, etc.), we have to replace the erosion operator by an erosion-like but error-tolerant operator.

Consider the rank filter that count how many points are in the window B centered at x and returns true if this value is greater or equal than a given $\lambda \in [\![\, 1,\, |B| \,]\!]$:

$$\kappa_B^\lambda(X) \;=\; \{\, x \in E \mid \sum_{b \in B} 1_{x-b \in X} \;\geq\; \lambda \,\}. \qquad (1)$$

We have $\delta_B(X) = \kappa_B^1(X)$ and $\varepsilon_B(X) = \kappa_B^{|B|}(X)$. An operator whose behavior is close to the one of the erosion is a rank filter with λ close to $|B|$; see [4].

To obtain a hit-or-miss-like operator that can extract staff chunks, we thus replace the erosion by a rank filter:

$$\kappa_{B_1}^{\alpha|B_1|}(X) \,\cap\, \kappa_{B_2}^{\beta|B_2|}(E \setminus X)$$

with α and β close to 1. Actually, this operator is the one we use to extract the thin lines that separate columns or paragraphs in document images; it is part of our C++ toolbox dedicated to document image processing and analysis [5]. The result of this operator is depicted in Figure 1b.

2.2. Regularization with a Horizontal Median Filter

Since the result of the first step is rather rough, we need to regularize it. To that aim, we rely on one of the most well-known non-linear image transformations: the median filter. Actually, this filter is related to mathematical morphology [6], and following Eq. 1, this filter is $\kappa_B^{|B|/2}$. With B being a horizontal segment, the regularizing effect can be observed in Figure 1c.

2.3. Horizontal Reconstruction

The result obtained after the two first steps does not perfectly match the contours of the original staff lines, since both the hit-or-miss and the median filter shift the object contours. The present step aims at extending horizontally the components obtained by the previous step, and at getting a binary image that envelops the original staff lines.

The geodesic dilation is defined by: $\delta \;=\; \delta_{\mathcal{N}(\mathbf{0}) \cup \{\mathbf{0}\}}$, where $\mathcal{N}$ is a neighborhood. The geodesic reconstruction by dilation of the set X, given the set Y such that $X \subseteq Y$, is defined by:

$$\mathcal{R}_Y^\delta(X) \;=\; \lim_{n \to \infty} \delta^n(X, Y)$$

where:

$$\delta^1(X,Y) \;=\; \delta(X) \,\cap\, Y,$$
$$\delta^{n+1}(X,Y) \;=\; \delta(\,\delta^n(X,Y)\,) \,\cap\, Y.$$

The intent of this operator is to iteratively dilate the marker X, while always keeping the result inside the mask Y (note that a fast implementation of geodesic reconstructions is given in [7]). In our case, the marker is the result of the previous step, whereas the mask is computed from the original image to enclose the staff lines. (Precisely the mask is obtained by $\varphi_{H'} \circ \delta_V \circ \gamma_H$, where the horizontal opening removes vertical objects , the vertical dilation enlarges the mask, and the final horizontal closing fills gaps and also enlarges the mask.)

The result is depicted in Figure 1d.

2.4. Cleaning

A cleaning step is then performed to correct little defects that can remain at that stage. (It mainly makes large parts of a same staff line do connect, when a small gap remains between them.) As it can be observed in Figure 1e, this step is usually a no-operation.

2.5. Line Selection

At the end of the previous step, we now have the staff lines. Unfortunately some false positive lines are also extracted; this is usually due to the presence of rather horizontal ties in the music score. Such a problem can be observed in the image excerpt of Figure 1e. To remove those spurious lines from the result of the previous step, we proceed to an analysis of the vertical periodicity of connected components. That allows to select the ones that correspond to staff lines. Figure 1e depicts in red the contour of the result of this step, superimposed over the input image. We can see that a spurious component has been properly removed.

Note that such an approach is classically used to estimate the staffline height and the staffspace height in many methods. In our case, the previous steps have considerably clean up the input image (while getting rid of non-staff elements) so its contents allow for a robust line selection. Incidentally we could now compute robust estimates for the staffline height and the staffspace height, but it is just of no use in our method.

2.6. Output Computation by a Local Median Filter

The result of the previous step can be considered as a binary mask, say K, and we have to remove the staff line pixels within this mask. Let us consider a filter ϕ dedicated to that job. With I being the input image, and O the output image, we have:

$$\forall\, p,\ \ O(p) \;=\; \begin{cases} \phi(I)(p) & \text{if } \ K(p) = \text{true} \\ I(p) & \text{otherwise} \end{cases}$$

for we do not want the filter ϕ to modify the input image outside the set K. Inside K, we have to keep the pixels of the set I if it belongs to a musical sign (other than a staff line); we can spot such an object because it crosses the staff line. As a consequence, $\phi = \kappa_V^{|V|/2}$ with V being a vertical segment, is a very good candidate. The result is depicted in Figure 1g.

3. RESULTS AND COMPARISON

Let us first describe the score image database [1], which contains a total of 12,000 base images. It is built from 20 music pages of different compositions transcribed by 50 different musicians, all adult so that they have their own characteristic handwriting style, and with as much heterogeneous background as possible. To those images, three levels of noise and two different kinds of mesh-based distortions have been applied.

method	H_1	H_2	M_1	M_2	L_1	L_2	*mean*
LRDE	**0.96**	**0.97**	**0.97**	**0.97**	**0.97**	**0.98**	**0.97**
NUASi-lin	0.92	0.94	0.93	0.95	0.93	0.95	*0.94*
NUASi-skel	0.92	0.93	0.92	0.93	0.93	0.93	*0.93*
Baseline	0.91	0.89	0.91	0.89	0.91	0.89	*0.90*
INESC	0.91	0.85	0.92	0.86	0.92	0.86	*0.89*
TAU	0.78	0.82	0.81	0.84	0.83	0.86	*0.82*
NUS	0.65	0.69	0.65	0.69	0.66	0.70	*0.67*
LRDE-gray	0.72	0.72	0.80	0.80	0.88	0.87	*0.80*
INESC-gray	0.39	0.36	0.39	0.36	0.39	0.37	*0.38*

Table 1: F-measure of the results w.r.t. to different methods (raws) and degradations (columns): H / M / L are respectively high / medium / low noise addition, and the subscript denotes one of the two different kinds of mesh-based distortions; our results are emphasized in bold faces.

The methods that took part of the ICDAR 2013 contest, and whose results are given in Table 1, are summarized in [2] and in [3]. We just recall here their bibliographical references: NUASi-lin and NUASi-skel [8] (in respectively Sec. II and Sec. III.D), TAU [9], Baseline [10], INESC [11, 12], and NUS [13]. The method presented in this present paper is labeled LRDE. Just note that two competitors have proposed a method to remove staff lines from gray-level images.

In the following, t, f, p and n stand respectively for "true", "false", "positive" and "negative"; for instance, tp is thus the number of "true positives". The results were evaluated by the F-measure: $F = 2PR\,/\,(P + R)$, where $P = tp\,/\,(tp + fp)$ is the precision and $R = tp\,/\,(tp + fn)$ is the recall. The accuracy is defined by $A = (tp + tn)\,/\,(tp + tn + fp + fn)$.

On the competition database, we obtained an average F-measure of 0.97 (precision 0.98 and recall 0.96), and an average accuracy of 0.998.

4. IMPLEMENTATION AND REPRODUCIBLE RESEARCH

We advocate *reproducible research* [14, 15], i.e., the idea that the ultimate product of academic research is not only the results presented in a scientific paper but also all the environment and data used to produce those results.

The CVC-MUSCIMA database of handwritten music score images [1] is available from

`http://www.cvc.uab.es/cvcmuscima`

it is free to be used for non-commercial research purpose only.

The method described in this present paper has been developped using our C++ image processing library, "Milena" [16, 17]. This library is part of the "Olena" Image Processing Platform, available as *Free Software*, and distributed under the conditions of the GNU General Public License (GPL) version 2. It can be downloaded from:

`http://olena.lrde.epita.fr` .

The full source code of our method (that requires Milena) is accessible from:

`http://www.lrde.epita.fr/wiki/Publications/geraud.14.icip`

For practionners that just want to test our method without manipulating code, an online demo can be run in a browser:

`http://olena.lrde.epita.fr/demos/staff_removal.php`

5. CONCLUSION

In this paper, we have presented a morphological filtering-based method to remove staff lines from music score images. We have shown that such an approach is a serious competitor of state-of-the-art methods. We want to emphasize that we provide the community with the source code corresponding to our method, and that we also offer an open source library containing a significant toolkit of mathematical morphology operators [17]. The reader who wants to learn more about mathematical morphology can refer to the two reference books from Jean Serra [18], or to the more recent couple of books intended to cover the largest part of this rich domain [19, 20].

Further qualitative discussions 1) on the keys that enable the results presented here as compared with state-of-the-art results, 2) on parameter selection, and 3) on the alternatives considered at each step of the process, are left for a forthcoming extended version of the present paper.

As a perspective, we want to improve our gray-level-based method (see "LRDE-gray" in Table 1), using morphological connected operators and some ideas from [21].

References

[1] A. Fornés, A. Dutta, A. Gordo, and J. Lladós, "CVC-MUSCIMA: a ground truth of handwritten music score images for writer identification and staff removal," *International Journal on Document Analysis and Recognition*, vol. 15, no. 3, pp. 243–251, 2012.

[2] V.C Kieu, Alicia Fornés, Muriel Visani, Nicholas Journet, and Anjan Dutta, "The ICDAR/GREC 2013 music scores competition on staff removal," in *Proceedings of the IAPR International Workshop on Graphics RECognition (GREC)*, Bethlehem, PA, USA, 2013.

[3] M. Visani, V.C Kieu, A. Fornés, and N. Journet, "The ICDAR 2013 music scores competition: Staff removal," in *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*, Washington, DC, USA, 2013, pp. 1407–1411.

[4] P. Soille, "On morphological operators based on rank filters," *Pattern Recognition*, vol. 35, no. 2, pp. 527–535, 2002.

[5] G. Lazzara, R. Levillain, T. Géraud, Y. Jacquelet, J. Marquegnies, and A. Crépin-Leblond, "The SCRIBO module of the Olena platform: a free software framework for document image analysis," in *Proc. of the Intl. Conf. on Document Analysis and Recognition (ICDAR)*, Beijing, China, 2011, pp. 252–258.

[6] H.J.A.M. Heijmans, "Self-dual morphological operators and filters," *Journal of Mathematical Imaging and Vision*, vol. 6, no. 1, pp. 15–36, 1996.

[7] T. Géraud, H. Talbot, and M. Van Droogenbroeck, *Mathematical Morphology—From Theory to Applications*, chapter 12: Algorithms for Mathematical Morphology, pp. 323–353, ISTE & Wiley, 2010.

[8] C. Dalitz, M. Droettboom, B. Pranzas, and I. Fujinaga, "A comparative study of staff removal algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 5, pp. 753–766, 2008.

[9] I. Fujinaga, *Adaptive Optical Music Recognition*, Ph.D. thesis, McGill University, USA, 1996.

[10] A. Dutta, U. Pal, A. Fornés, and J. Lladós, "An efficient staff removal approach from printed musical documents," in *Proc. of the Intl. Conf. on Pattern Recognition (ICPR)*, Istanbul, Turkey, 2010, pp. 1965–1968.

[11] J. dos Santos Cardoso, A. Capela, A. Rebelo, C. Guedes, and J. Pinto da Costa, "Staff detection with stable paths," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 6, pp. 1134–1139, 2009.

[12] J. Cardoso and A. Rebelo, "Robust staffline thickness and distance estimation in binary and gray-level music scores," in *Proc. of the Intl. Conf. on Pattern Recognition (ICPR)*, Istanbul, Turkey, 2010, pp. 1856–1859.

[13] B. Su, S. Lu, U. Pal, and C.L. Tan, "An effective staff detection and removal technique for musical documents," in *Proceedings of the IAPR International Workshop on Document Analysis Systems (DAS)*, Gold Coast, Queensland, Australia, March 2012, pp. 160–164.

[14] J.B. Buckheit and D.L. Donoho, "WaveLab and reproducible research," Tech. Rep. 474, Stanford University, Stanford CA 94305, USA, 1995.

[15] S. Fomel and J.F. Claerbout, "Guest editors' introduction: Reproducible research," *Computing in Science and Engineering*, vol. 11, no. 1, pp. 5–7, 2009.

[16] R. Levillain, T. Géraud, and L. Najman, "Milena: Write generic morphological algorithms once, run on many kinds of images," in *Mathematical Morphology and Its Applications to Signal and Image Processing—Proc. of the Intl. Symp. on Mathematical Morphology (ISMM)*. 2009, vol. 5720 of *Lecture Notes in Computer Science*, pp. 295–306, Springer.

[17] R. Levillain, T. Géraud, and L. Najman, "Why and how to design a generic and efficient image processing framework: The case of the Milena library," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2010, pp. 1941–1944, `http://olena.lrde.epita.fr`.

[18] J. Serra, *Image Analysis and Mathematical Morphology*, vol. 1 & vol. 2, Academic Press, London, 1982 & 1988.

[19] P. Soille, *Morphological Image Analysis: Principles and Applications*, Springer, 2 edition, 2004.

[20] L. Najman and H. Talbot, Eds., *Mathematical Morphology—From Theory to Applications*, ISTE & Wiley, 2010.

[21] G. Lazzara, T. Géraud, and R. Levillain, "Planting, growing, and pruning trees: Connected filters applied to document image analysis," in *Proceedings of the 11th IAPR International Workshop on Document Analysis Systems (DAS)*, Tours, Loire Valley, France, 2014.

[22] R. Levillain, T. Géraud, and L. Najman, "Writing reusable digital topology algorithms in a generic image processing framework," in *Applications of Discrete Geometry and Mathematical Morphology*. 2012, vol. 7346 of *LNCS*, pp. 140–153, Springer-Verlag.

[23] A. Rebelo, I. Fujinaga, F. Paszkiewicz, A. Marcal, C. Guedes, and J. Cardoso, "Optical music recognition: State-of-the-art and open issues," *International Journal of Multimedia Information Retrieval*, vol. 1, no. 3, pp. 173–190, 2012.