Approaches to handwritten conductor annotation extraction in musical scores

Eamonn Bell
Department of Music, Columbia University
New York
NY 10025, United States of America
epb2125@columbia.edu

Laurent Pugin
RISM (Switzerland)
Hallwylstrasse 15
CH-3000 Bern 6, Switzerland
laurent.pugin@rism-ch.org

ABSTRACT

Conductor copies of musical scores are typically rich in handwritten annotations. Ongoing archival efforts to digitize orchestral conductors' scores have made scanned copies of hundreds of these annotated scores available in digital formats.

The extraction of handwritten annotations from digitized printed documents is a difficult task for computer vision, with most approaches focusing on the extraction of handwritten text. However, conductors' annotation practices provide us with at least two affordances, which make the task more tractable in the musical domain.

First, many conductors opt to mark their scores using colored pencils, which contrast with the black and white print of sheet music. Consequently, we show promising results when using color separation techniques alone to recover handwritten annotations from conductors' scores.

We also compare annotated scores to unannotated copies and use a printed sheet music comparison tool to recover handwritten annotations as additions to the clean copy. We then investigate the use of both of these techniques in a combined method, which improves the results of the color separation technique.

These techniques are demonstrated using a sample of orchestral scores annotated by professional conductors of the New York Philharmonic. Handwritten annotation extraction in musical scores has applications to the systematic investigation of score annotation practices by performers, annotator attribution, and to the interactive presentation of annotated scores, which we briefly discuss.

Keywords

annotation extraction; image processing; color clustering; orchestral scores; conducting; image superimposition

1. INTRODUCTION

Handwritten annotations enrich documents with commentary and editorial revisions. Performers' annotations of musical scores indicate their musical preferences, and even authorial revisions when a performer is also the composer or editor of a score. Within the fields of computer vision and information extraction, there has been little emphasis on the identification and extraction of annotations from musical scores.

Early attempts at handwriting detection in scanned documents exploited the structural differences between handwritten and machine-printed characters to differentiate handwritten text annotations from printed text [8, 2]. Character-level probabilistic model have been used by [4] in the same task, while [3] achieved word-level annotation identification using Gabor filters.

The above techniques exploit text-specific features of handwriting and therefore cannot be used to extract non-text annotations. More general approaches show good results, extracting both handwritten and other basic geometric annotations using a probabilistic graphical model without relying on text-specific structural differences between handwriting and printed matter [6, 9].

Since most printed music scores of interest have been published at scale, we can often find a clean version of the score with which to employ image-comparison methods to perform annotation extraction. For instance, the IMSLP/Petrucci Music Library Project provides thousands of clean copies of public-domain scores for download. Our approach is in the spirit of [5], which used local arrangements of feature points to align both clean and annotated versions of a document in order to extract annotations.

Image processing techniques have been successfully applied to music source comparison in the Aruspix toolkit [7]. It enables the comparison of different copies of the same edition of printed music by image superimposition. Image processing is applied to de-skew, rotate, and resize the score images in order to align them. The differences between the two copies can be then extracted or highlighted.

Furthermore, many archival annotated scores have been digitized in full color. Therefore, information about the distribution of colors in digitized score images can be used as the input to classical image segmentation algorithms. We use this information to perform color separation by quantizing the color space of the annotated score images.

By using both image comparison and color separation we can achieve promising annotation extraction results without the use of traditional shift-invariant image features, text-specific features of handwritten text, or supervised learning techniques that require tagged examples of annotations.

Once the annotations have been extracted from a score image, they must be systematically associated with the under-

DLfM '16, August 12 2016, New York, USA

ISBN 978-1-4503-4751-8/16/08...$15.00

DOI: http://dx.doi.org/10.1145/2970044.2970053

lying score symbols in order to make sense of their meaning and function, which often depends on the musical context provided by the score. Annotations may be included in encoded versions of musical scores as SVG shapes according to the Music Encoding Initiative schema, though reliable heuristics for associating annotations to score symbols remain to be developed.

2. METHODOLOGY

2.1 Color separation using quantization

Handwritten annotations to scores are often in a color not represented in the printed sheet music. It is common for conductors to use colored pencils in the markup of scores for performance. We use color quantization to identify regions of an image array that correspond to colored annotations. Figure 1 summarizes this approach (Pipeline 1). A common method for computing a quantized color map from a source image (1) is to use an unsupervised clustering algorithm on the values of image pixels. We have used k-means clustering on a subsample of the marked score to train a classifier that partitions pixels by their value in the Lab perceptual color space into a number of visually similar clusters (2). The user identifies the clusters corresponding to the annotation colors, and the pixels corresponding to these clusters are selected and included in an image array that contains the annotated regions of the score image (3).

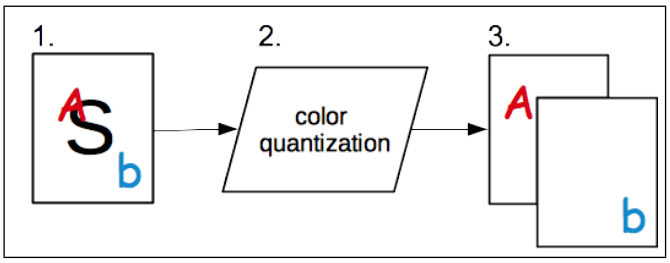


Figure 1: Extraction of colored annotations using color quantization on annotated score images (Pipeline 1). Here, and in subsequent figures, the black S represents regions of printed matter on the score image, while the colored letters represent the annotated regions on that image.

One disadvantage of this approach is that the k-means clustering algorithm is not deterministic. Certain initializations of the algorithm return quantizations that will not assign all the pixels of a colored annotation to the same cluster. One solution to this problem that we develop below is to bias the data upon which the k-means classifier is trained so that it is more likely to contain the pixel color values of annotated regions of the image. This improves the likelihood that the algorithm will converge on clusters that are representative of annotated regions of the score image. This approach, however, has limited applicabilty to scores that have been annotated using a pen or a pencil close in color to the color of printed ink on paper, or to grayscale scans of annotated scores.

2.2 Comparing marked and unmarked score images

We used Aruspix, a printed score comparison toolkit, to compare marked scores images to unmarked versions of their corresponding printed editions.

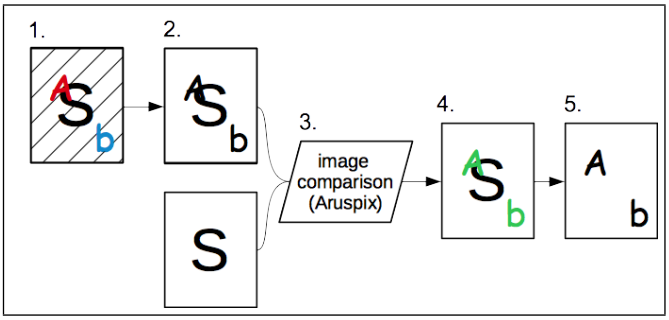


Figure 2: Comparing marked and unmarked scores to generate an "image diff" containing annotated regions (Pipeline 2).

This annotation extraction process (Pipeline 2) is summarized in Figure 2. Original color scans of marked scores (1) are pre-processed before grayscale conversion to adjust for variations in the neutral color of the paper. TIFFs of the same page of sheet music (2) are passed to Aruspix (3) as input along with alignment markers determined visually by the user. Aruspix returns an "image diff" that shows additions to the unmarked score in green (4); these additions to the score are stored in a black-and-white image array (5).

Since Aruspix currently only supports grayscale TIFFs at all stages of its internal processing pipeline, the "image diff" does not preserve color information if annotations have been made in color. However, this technique is applicable for the extraction of scores that have been annotated using a non-colored pen or pencil, or indeed to grayscale scans of annotated scores, in contradistinction to Pipeline 1.

2.3 Combined approach

The two approaches described above can be combined in Pipeline 3, as summarized in Figure 3. We start with the output of the image comparison pipeline, a black-and-white image array containing annotated regions of the original image (1). This array is converted to a mask (2) which reveals the annotated regions and hides the unannotated regions of the score image. The mask is then morphologically transformed by dilation (3). This increases the area revealed by the mask beyond the regions returned by the image comparison step. This step also improves the connectivity of these regions. The mask is applied (4) to the original image, and the color values of revealed pixels are used to train the k-means classifier that then performs color quantization (5). As before, clusters corresponding to annotation colors are filtered into separate image arrays.

This combined approach utilizes a mask defined by the results of the simple comparison pipeline to ensure that the input to the k-means clustering process contains the color values of as many pixels in annotated regions as possible. We aim to reduce the number of pixels representing blank paper or printed ink in the training data for the color quantization step. In turn, this increases the likelihood that the clustering algorithm will converge on clusters that represent the annotation colors in the original image, as opposed to elements of the printed score. The same mask may be reapplied at the end of the pipeline, obscuring any mislabeled regions of the whole score image that are not in the neigh-

borhood of annotated regions determined by the comparison step.

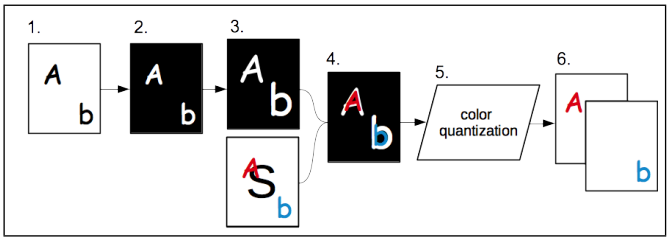


Figure 3: Combined approach, using results from image comparison pipeline to bias color quantization algorithm to improve convergence on colors representing annotations (Pipeline 3).

3. RESULTS

3.1 Results of extraction pipelines

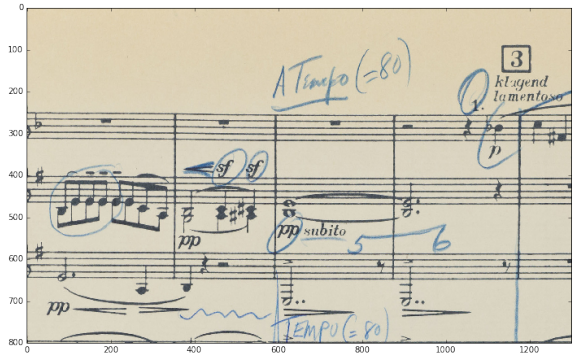
We show a cropped region of the result of applying the three processing pipelines to the example marked score in Figure 4 (a).[1] A post-processing filtering step was used to eliminate pixels that were unlikely to correspond to annotated regions. In order to eliminate artifacts resulting from the undesired extraction of vestigial remains of vertical and horizontal score elements and isolated single-pixel anomalies, three successive median filters with different window geometries were applied after each pipline was completed.

Figure 4 (b) shows the results of extraction by color separation, using k-means clustering ($k = 7$) in Lab perceptual color space (Pipeline 1). The quantizing classifier was trained on the color values of a random subsample of the pixels in the original image array. The resulting image shows the locations of color-quantized pixels corresponding to the annotation color (blue), manually selected by inspecting the color cluster centers. Figure 4 (c) shows the results of comparing marked and unmarked scores using Aruspix (Pipeline 2). Finally, Figure 4 (d) shows the results of the combined approach, performed by training the quantizing classifier only on pixels that appear in the neighborhood of additions to the score, in order to improve the likelihood that the k-means algorithm converges on clusters that represent annotations (Pipeline 3). Additionally, the dilated mask revealing the neighborhood of annotations was reapplied before post-processing, eliminating mislabeled pixels distant from suspected annotation regions.

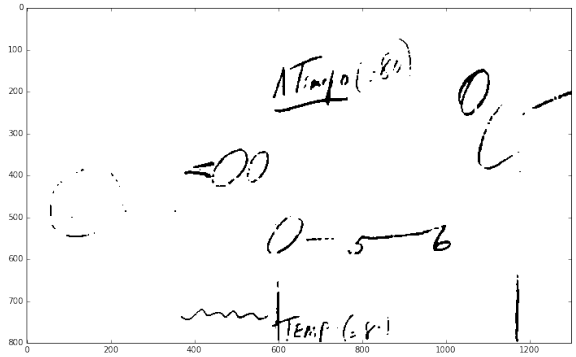
3.2 Discussion

While simple color separation (Pipeline 1) recovers parts of almost all of the original annotations, the results shown here indicate poor connectivity on large circular annotations and discontinuities in handwritten text. If the lighter parts of an annotation are sufficiently close to other non-annotation colors in the image, their color values may not be clustered with the color values of annotated regions of the image. The score image comparison pipeline (Pipeline 2) works well to address this issue, since any addition to the clean score will appear in the output of Aruspix's score

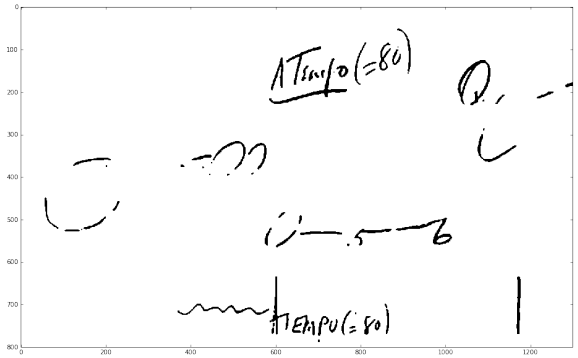
[1]The Python code used to implement each of these pipelines is available from the corresponding author (Bell), on request.



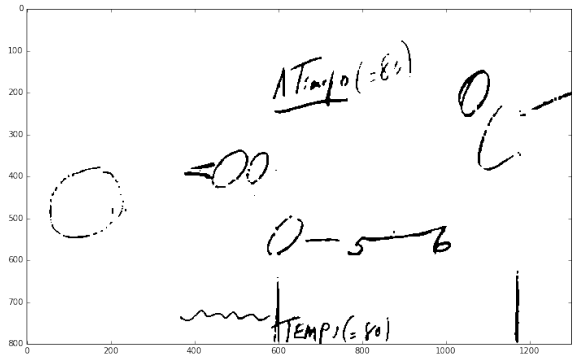
(a) Original score image (from Mahler, *Kindertotenlieder* (marked by Leonard Bernstein).



(b) Color separation (Pipeline 1).



(c) Image comparison (Pipeline 2).



(d) Combined approach (Pipeline 3).

Figure 4: Comparison of output of annotation extraction pipelines.

comparison tool. However, the circular annotations are still not fully connected, interrupted by erasures corresponding to staff lines, which appear in both annotated and clean copies of the printed music.

Connectivity of large annotations is most improved in the results of the combined pipeline (Pipeline 3), though the legibility of handwritten text is arguably superior in the results of straightforward score image comparison. The combined approach has the potential to remedy some of the issues of annotation extraction by color separation. Image comparison remains promising in general and, in particular, for cases where no color scan of the annotated score is available or where dark gray/black annotations are preferred by the annotator. In particular, we believe the combined approach demonstrates a novel joint application of image comparison techniques to improve the results of an unsupervised method of annotation extraction. In this case, the method improved upon was image segmentation by simple color quantization.

4. CONCLUSIONS

4.1 Prospects

Though the extraction processes described here produce promising results, a fully-automated, accurate workflow for score annotation extraction requires further work. For instance, during the score comparison step, Aruspix discards color information useful to annotation extraction. One possible improvement would be to modify the Aruspix algorithm in order to preserve this information and to make use of it throughout the comparison pipeline. Also, Aruspix's "image diff" must be manually transformed to match the geometry of the original score image for use in the masking step of the combined approach, though this issue concerns the interface to Aruspix, and not the underlying algorithms it utilizes. Improvements can be achieved by applying a staff removal algorithm to the annotated score image prior to extraction, such as those evaluated in [1]. It is expected that this would improve the connectivity of annotations that overlap staff lines. Ground truth for annotation extraction methods could be crowd sourced, with candidate annotation regions being determined using the score-comparison method. This can be used to perform useful cross-validation and parameter selection, along with automatic evaluation of extraction accuracy.

4.2 Applications

A wide range of applications of interest to digital libraries with annotated score holdings can be imagined that are based on the results above. Extracted annotations can be converted to vector graphics for integration into an encoded version of the score. The widely-supported Music Encoding Initiative schema supports the integration of SVG shapes within encoded music scores. This would make it possible for users to interactively and dynamically visualize different annotation sets that were made for the same underlying score, enabling interactive critical web editions of performance scores. Enriched scores could even lead to the use of annotations in the preparation of real or virtual performances that are musically informed by conductor annotations.

Dedicated visualization environments for the systematic investigation of annotations can also be imagined. For example, extracted annotations could be grouped by type using existing shape classification techniques. This would make it possible to display them grouped by shape, by score location, or by function.

Several scores exist that have been annotated by more than one conductor, sometimes with a single score being annotated by several conductors. Extracting the annotations is a first step towards eventually understanding the distinctive annotation practices of specific conductors, and towards testing authorship attribution hypotheses based on the content and structure of known-author score annotations.

5. ACKNOWLEDGMENTS

Mitchell Brodsky and Barbara Haws at the New York Philharmonic Archives for their technical support and encouragement. Original score image courtesy Leon Levy Digital Archive, New York Philharmonic.

6. REFERENCES

[1] C. Dalitz, M. Droettboom, B. Pranzas, and I. Fujinaga. A comparative study of staff removal algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5):735–766, 2008.

[2] K.-C. Fan, L.-S. Wang, and Y.-T. Tu. Classification of machine-printed and handwritten texts using character block layout variance. *Pattern Recognition*, 31(9):1275–1284, 1998.

[3] F. Farooq, K. Sridharan, and V. Govindaraju. Identifying handwritten text in mixed documents. In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 2, pages 1142–1145, 2006.

[4] J. K. Guo and M. Y. Ma. Separating handwritten material from machine printed text using hidden Markov models. In *Proceedings of the Sixth International Conference on Document Analysis and Recognition*, pages 439–443, 2001.

[5] T. Nakai, K. Kise, and M. Iwamura. A method of annotation extraction from paper documents using alignment based on local arrangements of feature points. In *Proceedings of the Ninth International Conference on Document Analysis and Recognition*, volume 1, pages 23–27. IEEE, 2007.

[6] X. Peng, S. Setlur, V. Govindaraju, and R. Sitaram. Handwritten text separation from annotated machine printed documents using Markov Random Fields. *International Journal on Document Analysis and Recognition (IJDAR)*, 16(1):1–16, Nov. 2011.

[7] L. Pugin. Aruspix: an automatic source-comparison system. In W. B. Hewlett and E. Selfridge-Field, editors, *Music Analysis East and West*, volume 14 of *Computing in Musicology*, pages 49–60. MIT Press, Cambridge, MA, 2006.

[8] S. Violante, R. Smith, and M. Reiss. A computationally efficient technique for discriminating between hand-written and printed text. In *IEE Colloquium on Document Image Processing and Multimedia Environments*, pages 17–1. IET, 1995.

[9] K. Zagoris, I. Pratikakis, A. Antonacopoulos, B. Gatos, and N. Papamarkos. Distinction between handwritten and machine-printed text based on the bag of visual words model. *Pattern Recognition*, 47(3):1051–1062, Mar. 2014.