Optical Music Recognition and Human-in-the-loop Computation

1nd Liang Chen
School of Informatics, Computing and Engineering
Indiana University Bloomington
Bloomington, USA
chen348@indiana.edu

2st Christopher Raphael
School of Informatics, Computing and Engineering
Indiana University Bloomington
Bloomington, USA
craphael@indiana.edu

***Abstract*—We present our work in optical music recognition in which we seek to transform scanned music notation images into symbolic representations. While music notation contains a small core of symbols and primitives composed in a rule-bound way, there are a great many common exceptions to these rules, as well as a heavy tail of rarer symbols. Since our goal is to create symbolic representations with accuracy near that of published music scores, we doubt the feasibility of fully-automatic recognition, opting instead for a human-interactive approach. We define a simple communication channel between the user and recognition engine, in which the user imposes pixel-level or model-level constraints, to improve our automatic OMR system.**

***Index Terms*—Optical Music Recognition, Human-in-the-loop Computation**

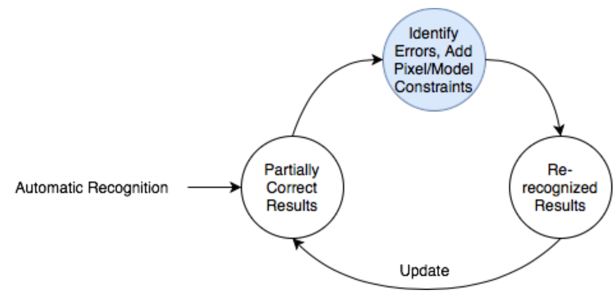

Fig. 1. System design of our human-interactive OMR system. The blue circle corresponds to the human work in the system.

I. Introduction

Optical Music Recognition (OMR) seeks to convert music score images into symbolic representations. Success with OMR would pave the way for large symbolic libraries containing all the world's public domain music, that could be instantly accessed, searched, transformed, and reformatted. Such libraries would provide a greatly improved experience for musicians through digital music stands; it would serve as the backbone for developing academic fields, such as computational musicology; and enable emerging applications fusing music and computation such as data-driven composition systems, musical accompaniment systems, and automatic music transcription.

OMR research dates back to the 1960s with mostly-disconnected approaches to many aspects of the problem [1]–[3] including several overviews [4], [5], and well-established commercial systems [6], [7]. In spite of these efforts there is still much to accomplish before the sought-after large-scale symbolic libraries will be in reach. The reason is simply that OMR is *hard*, constituting, in our view, one of the grand challenges of document recognition.

A. Challenges of Optical Music Recognition

Part of OMR's difficulty lies in the high degree of necessary recognition accuracy. The future's digital music stands will require accuracy at least as good as the familiar published hard-copy scores they will displace, otherwise this new technology will not be embraced.

In addition to the high bar regarding quality, OMR poses several significant technical challenges:

- The *two-dimensional* music layout is complex and hard to be analyzed, posing a harder problem than the text recognition.
- The music symbols are constructed from some basic grammatical rules, but these rules are sometimes violated.
- There is a *heavy tail* of standard music symbols. The symbols on the heavy tail are rare enough that their inclusion often results in more false positives than correct detections, yet they must be recognized.

B. Human-in-the-loop Computation

Given the demand for high accuracy and the technical challenges mentioned, we are skeptical that any fully automatic OMR approach will ever deal effectively with the wide range of situations encountered in real-life recognition scenarios. We formulate the challenge as one of *human-in-the-loop computation* instead of a fully automatic one, which fuses both human and machine abilities.

Our essential idea is to allow the user two axes of control over the recognition engine [8], [9]. In one axis the user chooses the *model* that can be used for a given recognition task, specifying both the exceptions to the symbols' construction rules, as well as the relevant variety of symbols to be used. In the other, the user labels misrecognized pixels with the correct primitive type, allowing the system to re-recognize subject to these user-imposed constraints. This approach results in a simple interface in which the user can provide a

Fig. 2. Interface of our human-interactive OMR system.

wealth of useful knowledge without needing to understand the inner-workings and representations of the system. Thus we effectively address the *communication* issue between human and recognition system.

II. Human-Interactive System, Performance and Conclusion

An overview of our system is illustrated in Fig. 1, where the human action is highlighted in the blue circle. The system is primed with the automatic symbol recognition, after which the system accepts the feedback from the human proofreaders and use it to automatically improve the recognition results. A screenshot of the system interface is shown in Fig. 2. In this example, the user is adding a pixel label for the system to re-recognize the missing small clef. The checking boxes and pull-down menu allows the user to change the model constraints.

We compare the performance of our system against one of the state-of-the-art commercial OMR systems *SmartScore*. While we cannot directly measure intermediate results, SmartScore's raw recognition accuracy appears to be significantly better than ours (at present). However, SmartScore takes less care with the correction of symbols, which holds the system back in an important way. That is, this system appears to be conceived primarily in terms of automatic recognition, with an afterthought that allows the user to correct individual errors. In contrast, our system was conceived as a human-in-the-loop system.

For both systems we concentrate on what happens after automatic recognition, focusing exclusively on the human effort necessary to correct the results. From the experimental results [10], we conclude that our system demonstrates performance that is competitive with SmartScore, in terms of both accuracy and efficiency. The data suggest that SmartScore was slightly more accurate while our system was slightly faster – users make different tradeoffs between these two objectives. In addition, it is worth noting that the data produced by our system also preserves more information about the precise construction and location of image symbols. While beyond the scope of our evaluation, such information can be integral to renotation [11], [12] approaches that leverage specific layout information from the original when creating newly formatted music notation.

References

[1] I. Fujinaga, "Optical music recognition system which learns," in *Proceedings of the SPIE - The International Society for Optical Engineering*, vol. 1785, 1993, pp. 210–17.

[2] G. E. Kopec, P. A. Chou, and D. A. Maltz, "Markov source model for printed music decoding," *Journal of Electronic Imaging*, vol. 5, no. 1, pp. 7–14, 1996.

[3] H. Fahmy and D. Blostein, "A graph-rewriting paradigm for discrete relaxation: Application to sheet-music recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 12(6), pp. 763–799, 1998.

[4] D. Blostein and H. S. Baird, "A critical survey of music image analysis," in *Structured Document Image Analysis*, 1992, pp. 405–434.

[5] A. Rebelo, G. Capela, and J. S. Cardoso, "Optical recognition of music symbols," *International Journal on Document Analysis and Recognition*, vol. 13, pp. 19–31, 2009.

[6] SmartScore, www.musitek.com, 1991.

[7] SharpEye, www.music-scanning.com/sharpeye.html, 2008.

[8] L. Chen and C. Raphael, "Human-directed optical music recognition," *Electronic Imaging*, vol. 2016, no. 17, pp. 1–9, 2016.

[9] L. Chen, E. Stolterman, and C. Raphael, "Human-interactive optical music recognition." in *ISMIR*, 2016, pp. 647–653.

[10] L. Chen, R. Jin, and C. Raphael, "Human-guided recognition of music score images," in *Proceedings of the 4th International Workshop on Digital Libraries for Musicology*. ACM, 2017, pp. 9–12.

[11] L. Chen, R. Jin, and C. Raphael, "Renotation from optical music recognition," in *International Conference on Mathematics and Computation in Music*, 2015, pp. 16–26.

[12] L. Chen and C. Raphael, "Renotation of optical music recognition data," in *SMC*, 2017.