Discussion Group Summary: Optical Music Recognition

Jorge Calvo-Zaragoza[1(✉)], Jan Hajič Jr.[2], and Alexander Pacha[3]

[1] PRHLT Research Center, Universitat Politècnica de València, Valencia, Spain
jcalvo@prhlt.upv.es
[2] Institute of Formal and Applied Linguistics, Charles University, Prague, Czech Republic
hajicj@ufal.mff.cuni.cz
[3] Institute of Visual Computing and Human-Centered Technology, Vienna, Austria
alexander.pacha@tuwien.ac.at

Abstract. This document summarizes the discussion of the interest group on Optical Music Recognition (OMR) that took place in the 12th IAPR International Workshop on Graphics Recognition, and presents the main conclusions drawn during the session: OMR should revisit how it describes itself, and the OMR community should intensify its collaboration both internally and with other stakeholders.

Keywords: Optical Music Recognition · Discussion group

1 Introduction

The 12th IAPR International Workshop on Graphics Recognition (GREC'17) hosted an interest group on Optical Music Recognition (OMR), a field of research that is concerned with computationally reading music notation in documents. OMR has been an active research field for decades, but so far it is (justifiably) known to "not work", at least not well enough for real-world use-cases.

The workshop was a unique opportunity for the field to reflect its state, as representatives of most active OMR research groups were present, thanks to the numerous workshop contributions related to the subject (10 out of 27 contributions at GREC'17). The attendees of the meeting were (in alphabetical order): Jorge Calvo-Zaragoza, Kwon-Young Choi, Jan Hajič jr., Jose M. Iñesta, Alexander Pacha, Zeyad Saleh, and Ké Zhang; Alicia Fornés spent some time in the discussion group as well.

The discussion uncovered broader systemic issues that hinder the progress of OMR towards usable systems, rather than just a lack of technical solutions. In this paper, we present the two most salient points that OMR needs to address:

– Revisiting how OMR is defined and described. This is necessary to design OMR systems that address actual needs and to accurately communicate the state of the art (Sect. 2).

J. Calvo-Zaragoza et al.—Equal contribution.

A. Fornés and B. Lamiroy (Eds.): GREC 2017, LNCS 11009, pp. 152–157, 2018.
https://doi.org/10.1007/978-3-030-02284-6_12

– Intensifying collaboration within the OMR community and with related fields, and making contributions to OMR more interoperable (Sect. 3).

2 Redefining OMR

The discussion group uncovered critical gaps in how OMR talks about itself. There is some intuitive understanding of what OMR is, and the standard pipeline [1,9,19] has been a helpful scaffold for decomposing OMR systems into manageable steps, but the field is starting to outgrow these foundations. For instance, with respect to methods, the traditional staff-line removal step [6] is no longer required in some systems before detecting musical symbols [10,17], and there are even end-to-end OMR systems that do away with most intermediate steps [5,21].

More importantly, OMR has been with few exceptions implicitly treated as a monolithic problem, with the accompanying assumption that it has a single goal, or at least a theoretically ideal output representation that OMR systems should produce (which may then be used for various purposes). As became evident while discussing a possible future OMR competition, this is not the case: it turns out there is only limited consensus on what are worthwhile objectives to compete in.

Existing OMR literature is of little help in this respect. The overwhelming majority of publications naturally focus on methods used to "solve OMR" [16, 19]. Some works are devoted to evaluation [2,7,13,18], and a paper by Byrd and Simonsen [3] analyses the various dimensions of OMR difficulty. However, there is no systematic treatment of the theoretical underpinnings of OMR: What is it actually trying to achieve? What is the internal structure of the field? There is an established taxonomy of OMR systems according to the inputs they process, but what is the taxonomy of OMR systems according to the *outputs* they should produce? Without a clearer idea of what OMR is expected to achieve, it is difficult to correctly evaluate sub-system improvements in the context of eventual OMR applications, and to communicate these advances — and their limitations — to stakeholders who are waiting for systems to be ready for their use-cases. Furthermore, such an analysis would make the problem of evaluating OMR more manageable as well.

The natural conclusion to this point of GREC'17 discussion is that an analysis of the field from the perspective of its goals and outputs should be performed and published.

3 Collaboration and Interoperability

Optical Music Recognition is by virtue of its domain interdisciplinary. Its motivations come not only from musicians and composers, but also from music libraries, musicology (especially its digital branch), and music information retrieval. On the other hand, none of these fields has the tools to provide solutions. For these,

OMR practitioners need to look to image processing, specifically document processing and pattern/graphics recognition, machine learning, and, in recent years, deep learning.

This dichotomy between communities that appreciate OMR results and those that can provide constructive feedback on OMR methods leads to the situation where OMR-related publications get scattered, and consequently their authors rarely meet in person. In this respect, the assembly of OMR researchers at the GREC'17 workshop was rather unique and was only possible because of a concerted effort of a member of the Program Committee, who reached out to active OMR researchers individually (since the field is small, this is manageable).

The International Society for Music Information Retrieval (ISMIR) conference has recently started attracting OMR contributions more naturally, since deep learning methods have proliferated in the music information retrieval community to the extent that deep learning-based OMR is a natural fit for the conference both in terms of applications and methods. However, despite these developments, the OMR community remains loose and its outputs are rarely inter-operable so that in effect it is difficult to actually build upon previous work.

Open-source software has, fortunately, become the (academic) norm. One has, e.g., Audiveris[1], the Pixel.js [20] and MUSCIMarker [11] data annotation tools, the pre-trained symbol detection models of Pacha et al. [17] in the Diva.js framework, and the CVC-MUSCIMA [8], HOMUS [4], and MUSCIMA++ datasets [12] that are available under liberal licenses, and of course the veritable Gamera open-source system [14]. However, there is not enough effort to ensure that data formats are inter-operable, evaluation procedures are shared among authors, and in general that the wheel does not get reinvented for every experiment.

A further critical missing piece for interoperability is the lack of a practical format for OMR-oriented structured representation of music notation. There is ongoing work in the MEI community[2], but it has limited reach. The MusicXML format, which is the *de facto* standard for music notation interchange, and MNX, its successor project led by the W3C Music Notation Community Group[3], are moving towards a broader standardization as well. However, both MEI and MusicXML/MNX are not very suitable for storing intermediate OMR information, as they mix together the graphical elements of music notation and the abstract musical objects that are encoded by them. There have been attempts to create a format tailored for OMR output [5,12,15], but so far none of them has actually become a standard.

A conclusion from this point of discussion is that an annual or biannual workshop centered on OMR that brings together its practitioners and stakeholders (and, to gain critical size, incorporating related "systems for reading music" — score following, cross-modal retrieval, and also music notation typesetting

[1] https://github.com/audiveris.
[2] http://music-encoding.org/.
[3] https://www.w3.org/community/music-notation.

software) should be organized to intensify collaboration within the OMR community, and build relationships to its stakeholders. This can be done at relatively low costs. Establishing personal contact and collaboration with digital musicology and digital libraries will also be necessary to sustain funding for OMR-related projects in the future.

4 Outlook

The GREC'17 workshop in Kyoto provided the OMR community with a unique opportunity to meet and discuss together the broad non-technical challenges the field is facing, which have unfortunately been somewhat neglected thus far. There are clear "action items" that the community should take upon itself to resolve:

- Revisit the way OMR talks about itself, specifically with the focus on a taxonomy of OMR systems according to their goals and outputs and accompanying evaluation metrics.
- Intensify collaboration within the OMR community and with related fields, preferably by creating a publication venue where stakeholders can naturally learn about each others' needs and use-cases and establish productive collaborations.
- Continue the trend of open-source software and data — make sure to reflect this principle e.g. when reviewing OMR publications.
- Improve music notation representations or at least provide conversion software between widely adopted representations such as MEI or MusicXML and OMR-specific formats.

The discussion group agreed that if this agenda for OMR is followed, the field will see a qualitative improvement that will ultimately benefit everyone involved: the OMR community itself, as well as the composers, musicians, musicologists, librarians, and other stakeholders who are waiting for reliable OMR systems that address their specific needs.

Acknowledgments. Jorge Calvo-Zaragoza acknowledges the support from the Spanish Ministerio de Economía, Industria y Competitividad through Juan de la Cierva - Formación grant (Ref. FJCI-2016-27873). Jan Hajič jr. acknowledges the support by the Czech Science Foundation grant no. P103/12/G084, Charles University Grant Agency grants 1444217 and 170217, and by SVV project 260 453.

References

1. Bainbridge, D., Bell, T.: A music notation construction engine for optical music recognition. Softw. Pract. Exp. **33**(2), 173–200 (2003)
2. Bellini, P., Bruno, I., Nesi, P.: Assessing optical music recognition tools. Comput. Music J. **31**(1), 68–93 (2007)

3. Byrd, D., Simonsen, J.G.: Towards a standard testbed for optical music recognition: definitions, metrics, and page images. J. New Music Res. **44**(3), 169–195 (2014)
4. Calvo-Zaragoza, J., Oncina, J.: Recognition of pen-based music notation: the HOMUS dataset. In: 22nd International Conference on Pattern Recognition, pp. 3038–3043 (2014)
5. Calvo-Zaragoza, J., Rizo, D.: End-to-end neural optical music recognition of monophonic scores. Appl. Sci. **8**(4), 606–629 (2018)
6. 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)
7. Droettboom, M., Fujinaga, I.: Symbol-level groundtruthing environment for OMR. In: Proceedings of the 5th International Conference on Music Information Retrieval, pp. 497–500 (2004)
8. Fornés, A., Dutta, A., Gordo, A., Lladós, J.: CVC-MUSCIMA: a ground truth of handwritten music score images for writer identification and staff removal. Int. J. Doc. Anal. Recogn. **15**(3), 243–251 (2012)
9. Fujinaga, I.: Exemplar-based learning in adaptive optical music recognition system. In: International Computer Music Conference, pp. 55–56 (1996)
10. Hajič Jr., J., Pecina, P.: Detecting Noteheads in Handwritten Scores with ConvNets and Bounding Box Regression. Computing Research Repository abs/1708.01806 (2017)
11. Hajič Jr., J., Pecina, P.: Groundtruthing (not only) music notation with musicmarker: a practical overview. In: 12th International Workshop on Graphics Recognition, pp. 47–48 (2017)
12. Hajič Jr., J., Pecina, P.: The MUSCIMA++ dataset for handwritten optical music recognition. In: 14th IAPR International Conference on Document Analysis and Recognition, pp. 39–46 (2017)
13. Hajič Jr., J., Novotný, J., Pecina, P., Pokorný, J.: Further steps towards a standard testbed for optical music recognition. In: Proceedings of the 17th International Society for Music Information Retrieval Conference, pp. 157–163. New York University (2016)
14. MacMillan, K., Droettboom, M., Fujinaga, I.: Gamera: optical music recognition in a new shell. In: Proceedings of the 2002 International Computer Music Conference (2002)
15. Miyao, H., Haralick, R.M.: Format of ground truth data used in the evaluation of the results of an optical music recognition system. In: IAPR workshop on document analysis systems, pp. 497–506 (2000)
16. Novotnỳ, J., Pokornỳ, J.: Introduction to optical music recognition: Overview and practical challenges. In: Necasky M., Moravec P., Pokorný, J. (eds.) Proceedings of the Dateso 2015 Annual International Workshop on DAtabases, TExts, Specifications and Objects, vol. 1343, pp. 65–76. CEUR-WS (2015)
17. Pacha, A., Choi, K.Y., Coüasnon, B., Ricquebourg, Y., Zanibbi, R., Eidenberger, H.: Handwritten music object detection: open issues and baseline results. In: 2018 13th IAPR Workshop on Document Analysis Systems (2018)
18. Padilla, V., Marsden, A., McLean, A., Ng, K.: Improving OMR for digital music libraries with multiple recognisers and multiple sources. In: 1st International Workshop on Digital Libraries for Musicology, pp. 1–8 (2014)
19. Rebelo, A., Fujinaga, I., Paszkiewicz, F., Marcal, A.R., Guedes, C., Cardoso, J.S.: Optical music recognition: state-of-the-art and open issues. Int. J. Multimedia Inf. Retrieval **1**(3), 173–190 (2012)

20. Saleh, Z., Zhang, K., Calvo-Zaragoza, J., Vigliensoni, G., Fujinaga, I.: Pixel.js: Web-based pixel classification correction platform for ground truth creation. In: 12th International Workshop on Graphics Recognition, pp. 39–40 (2017)
21. van der Wel, E., Ullrich, K.: Optical music recognition with convolutional sequence-to-sequence models. In: Proceedings of the 18th International Society for Music Information Retrieval Conference, pp. 731–737 (2017)