Expert Systems With Applications 89 (2017) 138–148

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Staff-line removal with selectional auto-encoders

Antonio-Javier Gallego, Jorge Calvo-Zaragoza*

Department of Software and Computing Systems University of Alicante, Carretera San Vicente del Raspeig s/n, 03690 Alicante, Spain

ARTICLE INFO

Article history:
Received 22 December 2016
Revised 13 May 2017
Accepted 4 July 2017
Available online 27 July 2017

Keywords:
Staff-line removal
Optical music recognition
Auto-encoders
Convolutional networks

ABSTRACT

Staff-line removal is an important preprocessing stage as regards most Optical Music Recognition systems. The common procedures employed to carry out this task involve image processing techniques. In contrast to these traditional methods, which are based on hand-engineered transformations, the problem can also be approached from a machine learning point of view if representative examples of the task are provided. We propose doing this through the use of a new approach involving auto-encoders, which select the appropriate features of an input feature set (Selectional Auto-Encoders). Within the context of the problem at hand, the model is trained to select those pixels of a given image that belong to a musical symbol, thus removing the lines of the staves. Our results show that the proposed technique is quite competitive and significantly outperforms the other state-of-art strategies considered, particularly when dealing with grayscale input images.

1. Introduction

Music is an important vehicle for cultural transmission, which is a key element as regards understanding the social, cultural and artistic trends of each period of history. A large number of musical documents have, therefore, been carefully preserved over the centuries and are scattered throughout cathedrals, libraries or historical archives (Fujinaga, Hankinson, & Cumming, 2014).

A significant effort has been made in recent decades to digitize these documents by means of scanners for their storage and distribution (Choudhury, Droetboom, DiLauro, Fujinaga, & Harrington, 2000). However, in order to make the music contained in the documents truly accessible, it is necessary for the images to be transcribed to a structured digital format that makes it possible to encode the content (notes, musical symbols, tonality, etc.) of the document. This also makes it possible to perform other interesting tasks, such as large-scale computational music analysis, search and retrieval by content, or transcription between different musical notations (Hankinson, Burgoyne, Vigliensoni, & Fujinaga, 2012).

The process of converting a scanned document into a musical structured digital format can be carried out manually by a user. The disadvantage is that it involves costs in terms of both resources and time. In addition, this process is especially tedious— because of the burdensome software for score edition— and very prone to introducing errors.

The research field known as Optical Music Recognition (OMR), which focuses on detecting and storing the musical content of a score from a scanned image (Raphael & Wang, 2011), has therefore been postulated as an important alternative that mitigates the aforementioned disadvantages of manual transcription. The objective of the OMR process is to import a scanned musical score and export its musical content to a machine-readable format 1), typically MusicXML or MEI.

The OMR task is similar to the recognition of text, typically known as Optical Character Recognition. However, unlike the text scenario in which words are analyzed sequentially with a single element to identify at each time instant, musical notation is considerably more difficult to recognize. This is principally owing to the possible existence of simultaneous matching elements, as in the case of polyphonic pieces with multiple notes that sound at the same time, thus resulting in several musical symbols being placed on the same interval. But it is also because of the presence of marks of expression, dynamics, articulations or even text to be sung in works with a vocal presence, among others.

Most current systems employ segmentation and classification approaches (Rebelo, Capela, & Cardoso, 2010; Wen, Rebelo, Zhang, & Cardoso, 2015). The first important obstacle that the OMR process must overcome is, therefore, the staff (or pentagram) lines: the set of five parallel lines on which musical symbols are located depending on their pitch. The staff-line removal stage is usually performed after the binarization of the document in the OMR

* Corresponding author.
E-mail addresses: jgallego@dlsi.ua.es (A.-J. Gallego), jcalvo@dlsi.ua.es (J. Calvo-Zaragoza).

http://dx.doi.org/10.1016/j.eswa.2017.07.002

(a) Example of input piece for an OMR system

(b) Symbolic representation of the piece

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

(a) Example of input score for an OMR system

(b) Input score after staff-line removal

Fig. 2. Example of a perfect staff-line removal process.

workflow (Rebelo & Cardoso, 2013). This binarization step helps to reduce the complexity of the problem, and it is advisable to apply strategies based on histogram analysis, connected components, or morphological operators.

Despite being necessary for musical readability, staff lines complicate the automatic detection and segmentation of symbols. Some specific works have taken advantage of specific features of printed and/or ancient notation to approach the problem of maintaining the staff lines (Calvo-Zaragoza, Barbancho, Tardón, & Barbancho, 2015; Ramirez & Ohya, 2014); however, the established OMR pipeline includes their detection and removal (Rebelo et al., 2012). This process must remove the staff lines while maintaining as much of the symbol information as possible (Fig. 2).

This paper proposes a framework with which to remove staff lines that is based on machine learning, that is, labeled examples can be used to train a model to perform the task. This allows using the same approach in a wide range of scores. We make use of a new approach of auto-encoder, which is trained to select only those pieces of the image that belong to musical symbols.

The remainder of the paper is structured as follows: Section 2 presents the background to staff detection and removal; Section 3 describes our approach with which to model the process; Section 4 contains the experimentation performed and the results obtained; and finally, the current work concludes in Section 5.

2. Background

This section presents the background to our approach. First, methods proposed for staff-line removal are introduced, after which the principles of auto-encoders are briefly described, given their importance as regards the present work.

2.1. Staff-line detection and removal

Staff-line removal has been an active research field for many years. A comprehensive review and comparison of the first attempts considered for this task can be consulted in the work of Dalitz, Droettboom, Pranzas, and Fujinaga (2008), who divided the staff-line removal strategies proposed until then into four categories: the Line Tracking (Bainbridge & Bell, 1997; Randriamahefa, Cocquerez, Fluhr, Pepin, & Philipp, 1993), Vector Field (Martin & Bellissant, 1991; Roach & Tatem, 1988), Runlength (Carter & Bacon, 1992; Fujinaga, 2005) and Skeleton (Ng, 2001) methods. However, given the interest in the task, many other methods have been proposed more recently.

Dos Santos Cardoso, Capela, Rebelo, Guedes, and Pinto da Costa (2009) proposed a method that considers the staff lines to be connecting paths between the two margins of the score. The score is then modeled as a graph so that staff detection is solved as a maximization problem. This strategy was later improved and extended to be used on grayscale scores (Rebelo & Cardoso, 2013).

Dutta, Pal, Fornes, and Llados (2010) developed a method that considers the staff line segment as a horizontal connection of vertical black runs with a uniform height. These segments are validated using neighboring properties before removing them.

Su, Lu, Pal, and Tan (2012) started by estimating the properties of the staves such as height and space. An attempt is then made to predict the direction of the lines and fit an approximate staff, which is subsequently adjusted.

Géraud (2014) developed a method that entails a series of morphological operators: first, a permissive hit-or-miss with a horizontal line pattern, followed by a horizontal median filter and a dilation operation. A binary mask is then obtained with a morphological closing. Finally, a vertical median filter is applied to the largest components of the mask. The procedure is directly applied to the image, which eventually removes staff lines.

Notwithstanding all the efforts made, the staff-removal stage is still inaccurate and often produces noise—staff lines not completely removed. Although it is possible to use more aggressive methods that minimize this noise, they may cause the 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 have led researchers to develop *ad-hoc* method for staff-line detection and removal. This results in methods that are not robust when applied to different types of document (from different eras, or with different notations or styles). In modern notation, a staff is composed by five black parallel lines over a white background. However, that is not always true in old music notation because (i) staff lines may appear with different ink colors, even closer to background color than to symbol color, (ii) handwritten staff lines are not totally straight, (iii) the thickness of the lines is irregular because of quill leakage, and (iv) the staff does may have less than five lines. In addition, lyrics in modern scores are always far enough from the staff, whereas there is much overlapping between music and lyrics in old notation, and so it could also hinder the staff-line removal process. Therefore, many of the assumptions for staff-line removal in modern music are not always fulfilled in different eras, thereby being extremely difficult