Expert Systems With Applications 72 (2017) 395–406

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Recognition of pen-based music notation with finite-state machines

Jorge Calvo-Zaragoza*, Jose Oncina

Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante, Carretera San Vicente del Raspeig s/n, 03690 Alicante, Spain

ARTICLE INFO

Article history:
Received 22 January 2016
Revised 31 August 2016
Accepted 17 October 2016
Available online 29 October 2016

Keywords:
Pen-based recognition
Optical music recognition
Finite-State machines

ABSTRACT

This work presents a statistical model to recognize pen-based music compositions using stroke recognition algorithms and finite-state machines. The series of strokes received as input is mapped onto a stochastic representation, which is combined with a formal language that describes musical symbols in terms of stroke primitives. Then, a Probabilistic Finite-State Automaton is obtained, which defines probabilities over the set of musical sequences. This model is eventually crossed with a semantic language to avoid sequences that does not make musical sense. Finally, a decoding strategy is applied in order to output a hypothesis about the musical sequence actually written. Comprehensive experimentation with several decoding algorithms, stroke similarity measures and probability density estimators are tested and evaluated following different metrics of interest. Results found have shown the goodness of the proposed model, obtaining competitive performances in all metrics and scenarios considered.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Despite several efforts to develop light and friendly software for music score edition, many composers still prefer to express their new musical compositions using a pen and paper. Once the artistic process is over, however, they resort to this kind of tools to transcribe the musical content into some kind of machine-readable format. Although this process is not always mandatory, it entails several benefits such as an easier storage, organization, distribution or reproduction of the music scores.

One profitable way in which to perform the whole process is by means of a pen-based music notation recognition system. These systems make use of an electronic pen, with which music symbols are drawn on a digital surface. The system collects user strokes and then processes them to recognize the composition. The goal is to present the score actually written to the user in the desired format. It should be noted that this task can be considered very similar to the Optical Character Recognition (OCR) task, for which pen-based research has been widely carried out (Liu, Yin, Wang, & Wang, 2013; Mondal, Bhattacharya, Parui, Das, & Roy, 2009; Plamondon & Srihari, 2000). Nevertheless, the complexity of musical notation in comparison to text leads to the need for specific developments (Bainbridge & Bell, 2001).

One straightforward approach that can be used to solve the task stated above is that of resorting to Optical Music Recognition (OMR) systems, whose purpose is to understand music scores from their image. That is, an image can be generated from pen strokes in order to make it pass through a conventional image-based system (offline recognition). Nevertheless, the performance of OMR systems is far from optimal, especially in the case of handwritten notation (Rebelo et al., 2012). Note that the main intention of a pen-based score composition system is to provide musicians with an interface that is as friendly as possible. The musicians should, therefore, be able to compose without having to pay attention to whether or not their handwriting is perfect. However, imperfect handwriting makes it even more difficult than usual to recognize the notation.

Fortunately, pen-based (or online) recognition provides new features that make the task very different to the offline case, some of which include:

- Staff lines: a staff is composed of five parallel lines, on which musical symbols are placed at different heights depending on their pitch. Staff-line detection and removal usually entails an obstacle that most offline OMR systems must overcome (Dalitz, Droettboom, Pranzas, & Fujinaga, 2008), since symbol detection and recognition are based on the accuracy of this step. Nevertheless, in a pen-based system the problem is insignificant because the lines in the staff are handled by the system itself and can be removed effortlessly.
- Segmentation: the input of a pen-based system is naturally segmented by pen strokes. Each stroke is easily detected by pen-down and pen-up actions over the digital surface. This makes it possible to avoid a lot of potential mistakes that may be caused by a bad segmentation in OMR systems.

* Corresponding author. Fax: +349 65 909326.
E-mail addresses: jcalvo@dlsi.ua.es (J. Calvo-Zaragoza), oncina@ua.es (J. Oncina).

http://dx.doi.org/10.1016/j.eswa.2016.10.041
0957-4174/© 2016 Elsevier Ltd. All rights reserved.

– Online data: drawing symbols in the pen-based scenario produces a time signal of coordinates indicating the path followed by the pen. Although the image of the score can be rebuilt from the strokes, online data is available to be used during recognition. It is known that this dynamic information is valuable for shape recognition (Kim & Sin, 2014).

All of the above features should lead to the development of specific pen-based algorithms that are able to improve the performance of current offline OMR systems.

This work proposes an approach with which to solve the problems involved in this task by using finite-state machines and dissimilarity measures between strokes. For a given input, the combination of these artifacts is able to produce a probabilistic model that defines the probability of each possible musical sequence. The use of decoding algorithms (for instance, searching for the most probable sequence) provides a hypothesis regarding the sequence actually written.

The use of finite-state machines makes it possible to build a system based on machine learning, signifying that the recognition is adaptable to any kind of handwriting style as long as training data is provided. Furthermore, these models establish an elegant statistical framework to solve the task, which is also supported by the vast background research conducted on finite-state machines (Mohri, Pereira, & Riley, 2002). It will be shown that this framework enables the easy incorporation of domain information — such as the fact that strokes are single independent units that can be grouped under the same label — and semantic constraints. Unlike that which occurs with similar frameworks such as hidden Markov models, this information prevents the use of pruning methods that may involve a loss of recognition accuracy and representativity (Rabiner, 1989).

The remainder of the paper is structured as follows: Section 2 presents some work related to pen-based music recognition; Section 3 provides details on the construction of the probabilistic model; Section 4 describes the experimentation carried out and the results obtained, and finally, Section 5 shows the main conclusions drawn and discusses some future work.

2. Background

Notwithstanding the benefits provided by pen-based music recognition systems for the composition of music, little attention has been paid to their development.

The first systems for the online recognition of musical scores were based on the use of simple gestures. This is the case of the *Presto* system, developed by Anstice, Bell, Cockburn, and Setchell (1996), which received as input short gestures that were generally mnemonic of the actual symbols they represented. These gestures were processed and translated into the actual musical symbols. The system was further improved in a later work (Ng, Bell, & Cockburn, 1998), including new capabilities based on a usability analysis of the previous approach. Poláček, Sporka, and Slavík (2009) developed a similar idea for use in low-resolution displays. These authors' alphabet consisted of gestures that were similar to conventional music notation, yet restricted to a more simple notation style. The main drawback of these approaches was that they did not provide a natural interface to musicians, who had to learn a new way of writing music.

More recently, many works have dealt with the problem of recognizing pen-based isolated musical symbols. George (2003) used the images generated by the digital pen to learn an Artificial Neural Network with which to recognize symbols. Her experimentation comprised 4 188 music symbols spread over 20 types of musical symbols from 25 different users. Although she completely ignored the time-signal information provided by the e-pen, she was able to obtain an accuracy of around 80%. Lee, Phon-Amnuaisuk, and Ting (2010b) proposed the use of hidden Markov models for the recognition of some of the most common musical symbols using different features of the shape drawn by the pen. The work was further extended in Lee, Phon-Amnuaisuk, and Ting (2010a), which considered other recognition schemes such as Naive Bayes or Markov models. These authors' results showed a virtually optimal performance. The conclusions drawn are, unfortunately, difficult to generalize since the experimentation only involved 400 samples, spread over 8 classes of symbols, from a single user. Calvo-Zaragoza and Oncina (2014), meanwhile, presented a free dataset of 15, 200 pen-based music symbols written by 100 different musicians, comprising 32 different music symbols. In addition to the dataset, they included an experimental baseline study taking into account several recognition algorithms. Their results reported that the time-signal information provided by the e-pen is able to outperform those classifiers that focus only on the shape drawn.

Nevertheless, while the recognition of isolated symbols might be of interest, the actual focus of this work is on the recognition of pen-based music sequences, which comprises a very different (and actually more difficult) challenge, and far fewer attempts have been made to deal with this issue.

Miyao and Maruyama (2004, 2007) proposed a system based on predefined stroke primitives (such as note-heads, lines, dots, etc.). The combination of the time-series data and image features were employed to recognize strokes using Support Vector Machines. Once the strokes had been classified, musical symbols were reconstructed following a heuristic approach that made use of a set of fixed restrictive rules in order to convert strokes into music symbols. The recognition rates were about 98 and 99% for strokes and music symbols, respectively. In spite of this performance, the approach did not enable users to write naturally, since the set of construction rules was very restrictive.

Macé, Éric Anquetil, and Couasnon (2005) proposed a generic approach for pen-based document recognition applied to music scores. This approach consisted of three modules: a stroke recognizer, a definition of the spatial structure of the document and a pen-based interaction framework. The recognition of the strokes was based on Neural Networks. The spatial structure of the document was modeled by means of *ad-hoc* Context-free Grammars, that took into account chronological and spatial information to define how and where strokes were expected by the system. This meant that any handwriting style that did not fulfill these constraints was not properly recognized. The interactivity framework allowed the user to validate or reject any output of the recognizer. Although no information regarding recognition rates was provided, the system was expected to recognize notes, accidentals, clefs, rests and barlines.

To the best of our knowledge, and supported by relevant reviews such as those by Rebelo et al. (2012) or Fornés and Sánchez (2014), no further research on the pen-based music recognition task has been carried out. As discussed above, the solutions proposed to date are not satisfactory from a user point of view since the only way in which to recognize symbols is by following the rules proposed by each system. These solutions have, therefore, forced users to adapt to the system style when what should be pursued is precisely the opposite.

All of the above reasons signify that there is still a need to develop an appropriate pen-based editing system for music scores. The main goals are to provide an ergonomic interface, which is indeed fulfilled with the use of the e-pen, and to provide an adaptive behavior that will allow musicians to use their own personal handwriting style. It is important to note that this question is not trivial. As an example, Table 1 shows some musical symbols written by different musicians, in which a great variability can be observed. This implies that recognition must be guided by a learning process

Download English Version:

https://daneshyari.com/en/article/4943443

Download Persian Version:

https://daneshyari.com/article/4943443

Daneshyari.com