Automatic conversion system of digital music score image based on image recognition algorithm

Yi Yao*
Xi'an Conservatory of Music, Xi'an 710000, Shaanxi, China

ABSTRACT

The development of target detection and recognition algorithms in the field of image processing has promoted the development of automatic image conversion systems for digital musical scores and related algorithms. This was crucial for the development of note-recognition coding and the conversion of musical scores into readable digital formats. The purpose of this paper is to study the automatic conversion system of digital music score image based on image recognition algorithm. The deep learning technology based on computer vision is used for musical score image recognition, and a digital musical score image automatic conversion system is designed and built. Harmonic peak method and confidence method are introduced, and their shortcomings are analyzed. Combined with the advantages of these two algorithms, an improved normal harmonic method is proposed, and the detailed ideas and steps of the algorithm are given. And the extracted fundamental frequency and time information are converted into pitch and time value combined with the basic knowledge of music theory. The recognition accuracy rate of the normal harmonic method proposed in this paper reaches more than 92%, which not only greatly improves the recognition accuracy, but also retains the strong anti-noise characteristics of the traditional harmonic peak method.

Keywords: Image recognition, digitized sheet music, sheet music image, automatic conversion

1. INTRODUCTION

Nowadays, with the development of computers, people gradually enter the field of information technology, the form of music begins to integrate with computers, and the pace of digitalization of floating music is gradually accelerated. As far as the current situation of music development is concerned, one of the difficulties facing people's music is the technicalization of music information[1, 2]. Among them, the process of shaping music is divided into two stages: composition and performance. Composition and performance are often separated, and computers can only accept code processing. So one has to break this barrier and let the computer recognize the image of the musical score[3, 4].

How to organically combine the art of music with computer science, that is, to automatically convert the images of paper scores to facilitate computer recognition and performance, has become an urgent problem to be solved[5]. Meng, F introduced a new notation called bootleg notation, which included the position of the points of the notes relative to the staff. MIDI representations can be converted to musical scores using Western music rules, and musical images can be converted to musical scores using classical computer techniques to recognize geometric shapes. Once the MIDI and phone images are converted into sheet music, the software can be used to simulate chords[6]. Brahmia, Z provides another way to integrate the content of two systems (OMR and AMT) to complete the integration process. Several experimental conditions were evaluated using monophonic musical compositions to assess the performance of individual transcription systems. In general, different methods are on average 40% more accurate than simple detection methods[7]. Because the music score image is complex, it not only contains the music score notes, but also the graphics that are easy to interfere with the recognition result, so there is no lack of research on the processing and recognition algorithm of the score image[8, 9].

This paper is an interdisciplinary research involving theoretical achievements in the following fields as prerequisites: its background involves the development and application of different common algorithms; its technology involves the establishment and basic usage rules of deep neural networks in the computer field; its data The analysis involves the role of traditional music theory in the statistical classification of music data, and the degree of automation of computer language functions in processing music data. Therefore, this paper is actually an exploratory combination of the research results in the above fields. This paper is not only an exploration of the development theory in the field of algorithmic

* yaoyipiano@qq.com

Third International Conference on Computer Science and Communication Technology (ICCSCT 2022)
edited by Yingfa Lu, Changbo Cheng, Proc. of SPIE Vol. 12506, 1250637
 · 0277-786X · doi: 10.1117/12.2662200

composition, but also a summary of the construction of a music deep learning composition system and the practical process of system operation.

2. RESEARCH ON AUTOMATIC CONVERSION SYSTEM OF DIGITAL MUSIC SCORE IMAGE BASED ON IMAGE RECOGNITION ALGORITHM

2.1 Basic knowledge of musical scores

Sound is produced by the vibration of objects. All the sounds that can be felt by human hearing are recorded as a set S, and each element (i.e., sound) in S can become a musical sound in a musical work[10]. Level, strength, length, and timbre are the four characteristics of sound. The sound level is determined by the number of times the sound vibrates; the duration of the sound determines the length of the sound, and the magnitude of the vibration determines the intensity, shape and nature of the sound. The branches of beats in music are called beats, and beats are marked as fractions. The tracker is used to indicate the number of singles in each scale, the denominator is used to indicate the recorded duration of the singles, and the horizontal line in the punctuation is replaced by the third line on the staff.

2.2 Image preprocessing

2.2.1 Binarization. Fractional images are usually grayscale images and contain no color channels. For a typical standard score image, the foreground color pixels (lines and annotations) are generally pure black, and the background color pixels are pure white. Therefore, the foreground and background must be separated, and the reproduction of the score image can maximize the separation.

2.2.2 Noise Reduction. The adaptive filter obviously increases the complexity of the algorithm, but the processing effect is very good, and a clearer picture can be obtained. The original score image is usually superimposed with salt and pepper noise, so this paper adopts an intermediate filter as the noise reduction method.

2.3 Image recognition algorithm

2.3.1 Harmonic Peak Method. The Harmonic Peak method is a standard algorithm based on velocity modulation. It shows the relationship between the frequency and amplitude of a signal and is therefore widely used to calculate the frequency spectrum of a signal. The best matching method means that the peak with the largest amplitude in the spectrum corresponds to the fundamental frequency of the audio signal, so its frequency is considered to be the fundamental frequency. The biggest advantage of this method is that it is very simple and requires very little time and space. In actual technical work, especially in the form of musical instruments, especially in some bass areas, the fundamental frequency of the spectrum is not necessarily high.

2.3.2 Confidence Method. In the case where the peak amplitude of the harmonic is higher than the fundamental wave, the confidence-based optimization algorithm considers that the component with the largest peak amplitude should be the fundamental wave or the *n*th (usually no more than 5) harmonics as the candidate fundamental frequency using 1 to 5 and then add the inverse of the nth harmonic. The confidence-based algorithm solves the problem of the maximum peak-width component harmony to a certain extent. When dealing with low-frequency sound waves, the amplitude of the fundamental component in the spectrogram is usually very low or even no, the higher harmonic components are more and the amplitude is larger.

3. INVESTIGATION AND RESEARCH ON AUTOMATIC CONVERSION SYSTEM OF DIGITAL MUSIC SCORE IMAGE BASED ON IMAGE RECOGNITION ALGORITHM

3.1 System development environment

OpenCV is a frequently used computer vision library, which contains a large number of excellent algorithms, including target detection, target tracking, contour detection, OCR algorithm, distortion correction, image denoising, image enhancement and image binarization processing etc. The library is very easy to use, and you only need to call a simple interface to complete the corresponding functions.

The cv2.imread (img, 1) function can read the picture; “img” means the original picture to be read; 1 means reading the color picture; 0 means reading the grayscale picture. The cv2.fastNlMeansDenoising (img, None, 10, 7, 21) function denoises the image; 10, 7 and 21 represent some hyperparameters that will be used in the algorithm. The cv2.imwrite (save_name, img) function can save the image function; save_name represents the name of the image to be saved; img

represents the original image input. cv2.getTextSize (text, font, fontScale=0.7, thickness=1) function can display text information on the rectangular box; text represents the text to be displayed; font represents the displayed text font; thickness represents the thickness of the text.

3.2 Experimental setup

The identified experimental environment is built with VisualStudio2010+OpenCv2.4.8 under Windows. In the experiment, the recording of 88 piano keys with a duration of 2 seconds recorded in a basically noise-free environment was selected as the test sample from 40dB to 70dB.

3.3 Improved normal harmonic method

In this paper, an improved normal harmonic method is used to extract the fundamental frequency. First, we use the discrete first and second derivatives to find the top n maximum points x1, x2, ..., x*n*-1, x*n* with higher peaks in the spectrogram as candidate frequencies to construct a confidence function *h*(*xi*) to reflect the possibility that x1 is the fundamental frequency:

$$h(x_i)=\sum_{k=1}^{n} g(x_k)t(x_k,x_i) \quad (i=1,2,3,....,n) \tag{1}$$

Among them, $g(x_k)$ is the energy value corresponding to the candidate frequency *xk*; $t(x_k, x_i)$ represents the proximity of x_k to an integer multiple of xi. For any candidate fundamental frequency x_i, $t(x_k,x_i)$ is defined as follows:

$$t(x_k,x_i)=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(\theta_k)^2}{2\sigma^2}} \quad (k=1,2,3,...,n) \tag{2}$$

In equation (1), the confidence $h(x_i)$ of the candidate fundamental frequency *xi* is determined by the energy value $g(x_k)$ of all candidate fundamental frequencies and the proximity degree $t(x_k, x_i)$ of the candidate frequency to an integer multiple of *xi* The sum of the products of. For a given value of k, the energy value $g(x_k)$ of the candidate fundamental frequency x_k is a constant. At this time, only the closer x_k is to an integer multiple of xi, the larger $t(x_k, x_i)$ is, and the more likely x_k is Nth harmonic frequency of the candidate fundamental frequency x_i. Therefore, $h(x_i)$ can effectively reflect the possibility that x_i is the fundamental frequency.

Considering that there is a certain error in the ratio of frequency multiplication and fundamental frequency after FFT, the rate at which the curve decreases near a positive integer can be adjusted by changing the scale parameter σ in equation (2). From the properties of normal distribution, increasing the value of σ can make the curve smoother to increase the fault tolerance rate of the curve, and conversely, it can make the curve steeper to reduce the fault tolerance rate of the curve. If x_i is a multiplier, then x_k/x_i is very likely to be close to a positive integer, and $t(x_k, x_i)$ is very high; if x_i is noise, then x_k/x_i is usually far away from a positive integer, $t(x_k, x_i)$ is very low, In this way, most of the influence of noise on the fundamental frequency confidence calculation is reduced.

4. ANALYSIS AND RESEARCH ON AUTOMATIC CONVERSION SYSTEM OF DIGITAL MUSIC SCORE IMAGE BASED ON IMAGE RECOGNITION ALGORITHM

4.1 Automatic conversion system of digital music score images

The schematic diagram of the image processing algorithm is shown in Figure 1. The whole algorithm includes four main parts: image processing, operator detection, detection and classification, and reconstruction. The purpose of image processing is to remove some information that interferes with the image, so as to facilitate the implementation of subsequent algorithms, including image enhancement, image fusion, image processing and image processing. The main purpose of sound detection is to accurately identify the specific position of the note in the score, since all the notes are drawn on a line, identifying the correct note in the score is a fundamental task.

4.2 Music recognition

Its recognition accuracy is shown in Figure 2. In the figure, the abscissa represents the signal-to-noise ratio of the added Gaussian noise, and the ordinate represents the accuracy of fundamental frequency identification of different algorithms.

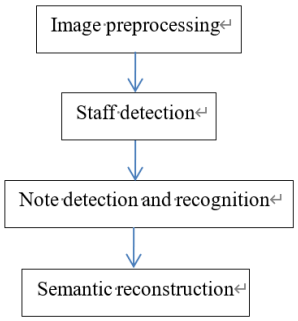

Figure 1. System workflow framework diagram.

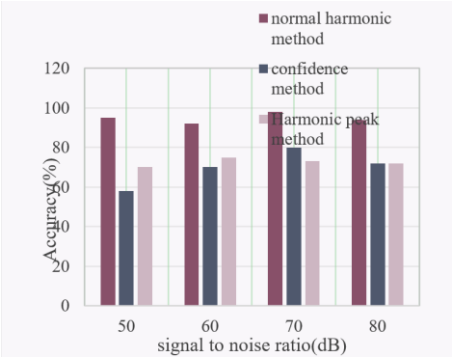

Figure 2. Comparison of accuracy rates of fundamental frequency extraction algorithms for piano 88-key recording.

In all cases, the normal harmonic method has the highest recognition accuracy, reaching an accuracy of more than 92%, and the recognition rate of this method does not decrease significantly when the noise gradually increases, reflecting that the method is not sensitive to noise shown in Table 1. Under the condition of high signal-to-noise ratio, the confidence method has a recognition accuracy of up to 80%, which is higher than 75% of the harmonic peak method. However, with the increase of noise, after the signal-to-noise ratio drops below 60 dB, the accuracy of the confidence method drops sharply. When the signal-to-noise ratio is 50 dB, the recognition accuracy of the confidence method is only 58%, while the harmonic peak. The accuracy of the method remains stable at around 70%.

Table 1. Comparison of recognition accuracy of three algorithms.

Signal to noise ratio (dB)	Normal harmonic method	Confidence method	Harmonic peak method
50	95	58	70
60	92	70	75
70	98	80	73
80	94	72	72

5. CONCLUSIONS

In recent years, with the rapid development of computer technology, the pace of digitalization of music information is getting faster and faster. This paper mainly realizes the automatic conversion system of digital music score image based on the image recognition algorithm. The main work completed is as follows. The overall process of the music score image recognition system is designed; the hardware equipment required to build the system is selected according to the comprehensive analysis of the system; the Windows system with excellent openness and high operating efficiency is selected as the system software environment; the OpenCv open source visual library is used to analyze the collected data. The digital music score image is processed by size transformation, brightness transformation and so on. The system still has many shortcomings and needs to be improved. At present, the system only recognizes the notes in the musical score, and is not designed to recognize complex musical score information such as clefs, diacritics, and rests. Therefore, the function of the system is not perfect, and the recognition ability of complex musical scores is still lacking, which needs further research.

REFERENCES

[1] Mueller, M., Arzt, A., Balke, S., Dorfer, M. and Widmer, G., “Cross-modal music retrieval and applications: An overview of key methodologies,” IEEE Signal Processing Magazine, 36(1), 52-62(2018).

[2] Dar, S. A. and Madhusudhan, M., “Digital nomadism: Students experience of using mobile devices in delhi metro,” Library Hi Tech News, 35(7), 5-10(2018).

[3] Bostoen, F. and Vanherpe, J., "Competition law in the digitized music industry: the winners take it all - but should they?," Competition Policy International, (2), 30-38(2021).
[4] Tsai, T. J., Yang, D., Shan, M., Tanprasert, T. and Jenrungrot, T., "Using cell phone pictures of sheet music to retrieve midi passages," IEEE Transactions on Multimedia, 22(12), 3115-3127(2020).
[5] Mathew, P., Vijayakumar, R., Kuriakose, A. T., Sunny, J. and Ramani, B. V., "Optical music recognition using image processing and machine learning," International Journal of Computer Sciences and Engineering, 6(10), 18-23(2018).
[6] Meng, F., Lu, T. and Li, F., "Stabilization of solvent to α-sheet structure and conversion between α-sheet and β-sheet in the fibrillation process of amyloid peptide," The Journal of Physical Chemistry. B, 123(45), 9576-9583(2019).
[7] Brahmia, Z., Grandi, F. and Bouaziz, R., "Conversion of xml schema design styles with style volution," International Journal of Web Information Systems, 16(1), 23-64(2020).
[8] Gomez-Alanis, A., Gonzalez-Lopez, J. A., Dubagunta, S. P., Peinado, A. M. and Magimai-Doss, M., "On joint optimization of automatic speaker verification and anti-spoofing in the embedding space," IEEE Transactions on Information Forensics and Security, 16, 1579-1593(2020).
[9] Khairnar, K. and Khan, S., "Automatic early leaf spot disease segmentation on cotton plant leaf," International Journal of Recent Technology and Engineering, 9(2), 2277-3878(2020).
[10] Chung, M. J., Hirose, T., Ono, T. and Chen, P. H., "A 115× conversion-ratio thermoelectric energy-harvesting battery charger for the internet of things," IEEE Transactions on Circuits and Systems I: Regular Papers, 67(11), 1-12(2020).