Research on Virtual Piano Based on Computer Binocular Stereo Vision

Lei Zhao*
School of Music and Dance, Zaozhuang University 277160

*Corresponding author e-mail: Zhaolei@163.com

Abstract. The binocular stereo vision technology emerges along with the development of digital technology, which has been widely employed in virtual piano to effectively remedy the user-unfriendliness and unnaturalness of the keyboard and mouse used as the key. This paper discusses how binocular stereo vision technology can be made good use to realize the function of the piano so that a comfortable real-time human-computer interaction can be completed.

Keywords: Binocular Stereo Vision, Virtual Piano, Function

1. Introduction
The blooming of computer technology promotes the advanced technology concerning human-machine interaction, in which field computer vision has deemed as the breaking point to make computers more automatic and smarter. Its application has made daily life more convenient, also recreational life. In recent years, along with increasing spiritual needs, more and more people hope to learn the piano, which is a keyboard instrument that originated in Western classical music and is considered as the king of musical instruments for its beautiful timbre and wide range.

However, it is challenging to play the piano wherever and whenever one would like since it is bulky and inconvenient to carry. In this case, the virtual piano came into being and is becoming more and more popular (see Figure 1)[1].
At present, the popular virtual piano on the market also belongs to a kind of electronic piano in a broad sense, which is generally a simple software developed based on a specific computer language. The user can operate a particular key by clicking the mouse or pressing the keyboard, thereby triggering the audio file corresponding to the key, making the key ring, and controlling the playing sound by adjusting the volume button on the interface. However, in terms of operation methods, this type of virtual piano does not have a real key carrier. It only has a traditional input device, which is in accordance with people's habits of using a mouse or a keyboard and thus cannot meet users’ requirements for piano playing. Therefore, the research on the development of a new type of virtual piano is of considerable significance, which can overcome the deficiencies traditional pianos have, and also improve the existing virtual pianos[2].

The increasing popularity of smart handheld devices has made virtual keyboards develop rapidly. The binocular virtual keyboard removes the infrared light source in the monocular device, which uses two cameras to track the finger in real-time and calculate the position of the finger in the three-dimensional world to determine the target key and implement character input. Compared with the expected virtual piano, the virtual keyboard only performs different functions. The methods of determining the keyboard area and identifying the target keys still provide a lot of ideas for this research.

2. Definitions of binocular stereo vision and virtual key

2.1. Definition of binocular stereo vision

Based on computer vision technology, binocular stereo vision technology intends to use a camera to imitate human eyes, and then uses the two-dimensional image acquired by the camera to reconstruct the three-dimensional shape of the object in order to accurately locate the position of the object and determine it as the goal of a computer reconstruction of object. This technology is widely used, such
as in 3D movies. Binocular stereo vision has the advantages of simple and reliable operation and strong operability and is widely accepted and applied by people[3].

2.2. Definition of virtual key
Virtual reality technology ushered in a new era of human-computer interaction. At present, virtual keyboard mainly uses a single camera and binocular-recognition-based image tracking.

Virtualization of the keyboard through the monocular is to detect the state of the finger through an infrared light source to determine the target key. The virtual keyboard formed by this method has the characteristics of simple operation, single structure, and low accuracy. And when using the virtual keyboard, it will cause certain harm to people’s bodies. On the contrary, the virtual keyboard based on binocular vision technology can quickly locate the target position and has little influence on the external environment.

3. The overall structure of the virtual piano system
The hardware of the virtual piano system based on binocular stereo vision is mainly composed of three parts: a personal computer, two CMOS cameras with the same functional parameters, and a homemade paper piano keyboard. This system functions as the following Figure (see Figure 2). In practice, firstly, the target key pressed by the finger, is identified by combining the depth of the fingertip and its image coordinates. Next, three boolean quantities are used to characterize the key pressing and releasing modes, in which process the virtual piano playing function is realized. Then, the errors of the virtual piano system are analyzed from the offline modelling of key positions and binocular stereo vision ranging. In the end, the accuracy of playing the keys verifies that the implementation of virtual piano based on binocular stereo vision technology is feasible.

The detailed description of the essential moves of this system will be discussed in the following parts[4, 5].

![Figure 2. Overall process of the system](image)

4. Offline modelling of key positions
The key positions will be confirmed through offline modelling. And then a model will be made and saved for use. Firstly, the image will be preprocessed. Preprocessing the image is to convert the original image into a high-quality binary image through some basic operations in preparation for contour search. Due to various factors, such as devices and lights, the image collected by the camera will contain a lot of noises. If edge detection is performed directly, a lot of false edges will be
obtained. In order to eliminate noise interference, the captured grayscale image is filtered. After comparison, it is found that the noise reduction effect of Gaussian filtering is the most obvious in the environment of the system. After the image is Gaussian filtered, canny edge detection algorithm is used to detect the edge of the white key, and the grayscale image is further converted into an edge binary image. However, because edge breakpoints will be generated after edge detection, which is not conducive to extracting complete white key contours, morphological dilation processing is used to connect contour breakpoints[6].

5. The ranging model and camera positioning
Binocular vision ranging is to use two cameras with the same parameters to obtain two images of the same object from different viewpoints, then measures the parallax of the object in the two images, and later uses the principle of binocular vision imaging to calculate the distance from the target point to the camera’s optical center plane.

To obtain the depth of the fingertips, that is, the distance from fingertips to the light center plane of the left and right cameras, the exact values of the focal length F and optical center distance T of the two camera pixels must be known. First, the left and right cameras are individually calibrated to obtain their internal and external parameters, and the pixel focal length F is obtained from the internal parameters. Then, the external parameters of the two cameras are used to perform stereo calibration on the binocular vision system, thereby obtaining the optical center distance T of the two cameras.

6. Image segmentation and fingertip detection
The purpose of image segmentation is to separate the regions of interest in the image. In computer vision, image segmentation is significant, through which objects that need to be detected or tracked can be obtained, which will significantly reduce the amount of calculation in subsequent parts.

After the foreground image of the finger is obtained, the position of the fingertip must be accurately determined, and the difference between the fingertip coordinates in the left and right cameras in the horizontal direction can be obtained to calculate the three-dimensional depth of the fingertip, and then determine whether the finger presses the key or lift off the keys. At present, commonly used algorithms for fingertip detection include the Hough transform, the contour analysis, and special marking.

7. Summary
This paper mainly explores the development of virtual piano based on computer binocular stereo vision, including research background and significance, overview and detailed discussion of the overall structure of the virtual piano system. In the detailed description, key moves such as offline modelling of key positions, the ranging model and camera positioning, image segmentation and fingertip detection and implementation of the virtual piano are presented. Above all, this research provides a model for the future exploration and development of the virtual piano.

References
[1] Chitnis O, Borkar A, Donde Y. Virtual Piano Using Image Processing[J]. 2017.
[2] Kang S, Kim J, Yoon S. Virtual Piano using Computer Vision[J]. arXiv preprint
arXiv:1910.12539, 2019.

[3] Patel A, Satpute A, Pattani M, et al. Virtual Piano[C]//Proceedings of International Conference on Wireless Communication. Springer, Singapore, 2018: 265-272.

[4] Qiao W, Wei R, Zhao S, et al. A real-time virtual piano based on gesture capture data[C]//2017 12th International Conference on Computer Science and Education (ICCSE). IEEE, 2017: 740-743.

[5] Sawamura H, Gillebert C R, Todd J T, et al. Binocular stereo acuity affects monocular three-dimensional shape perception in patients with strabismus[J]. British Journal of Ophthalmology, 2018, 102(10): 1413-1418.

[6] Yang L, Wang B, Zhang R, et al. Analysis on location accuracy for the binocular stereo vision system[J]. IEEE Photonics Journal, 2017, 10(1): 1-16.