

Design and Realization of Kinect Somatosensory Fitting System Based on WEB

LI Guangsong

Department of Information Engineering
Guangdong Polytechnic
Foshan, China
e-mail: ivan_lgs@163.com

Abstract—*In order to solve the problems of try on the clothes and returned purchase in electronic commerce, it is using Kinect sensor to match clothing and the human body, through the simulation experiment to build the fitting room by downloading and installing the WEB client. It firstly uses the Kinect sensor to acquire the depth image information, and then by hand analysis module to obtain palm position, and finally through point matching algorithm, it matches human bones with clothes to get the network fitting effect. Experiments show that, this method has higher recognition rate and shorter response time. Method the system provides has strong stability and real-time, and can meet the requirements of network somatosensory fitting.*

Keywords- Kinect; somatosensory; WEB; point matching; fitting system

I. INTRODUCTION

With the development of e-commerce, including clothing, online shopping has become common. But it is unable to try clothing like in traditional entity shop, so it is difficult to determine the own suitability. How to provide the time-saving, labor-saving and worry-saving virtual fitting method is bringing more and more attention from home and abroad. Somatosensory fitting is using hand position recognition as a way of human-computer interaction, and as long as the customers stand in the front of somatosensory fitting system, it will be free to choose their own clothes without changing real clothes [1].

With the development of camera technology, the cameras which can shoot 3D information come into the line of vision of people [2]. The TOF (time of flight) can obtain the depth information by calculating the light flight time to measure the photographed object. Compared with the traditional 2D camera, TOF camera is easy to distinguish the foreground and background in the scene, and has unique advantages in target recognition and tracking. But the TOF camera in the application has shortcoming of higher price and lower resolution [3]. Microsoft launched 3D camera Kinect for Xbox 360 somatosensory peripheral in 2010, as shown in Fig .1, which is using a structured light coding technology to obtain the depth information of images. The imaging part of the Kinect includes a GRB camera, an infrared camera and an infrared transmitter. The infrared emitter can emit near-infrared laser, and when the laser beam reaches the rough objects, the diffraction spots with high degree of randomness will be formed, known as laser speckle [4-5]. Through analysis of the spatial distance obtained by speckle pattern, image processor generates the depth image, and uses the USB2.0 protocol to transmit to the PC terminal. Compared with the TOF camera, Kinect is cheaper, and the resolution shooting can reach pixel level. Due to the use of the image processor chip, it does not need additional computation in the PC terminal, and can make the PC terminal low configuration to achieve real-time requirements.

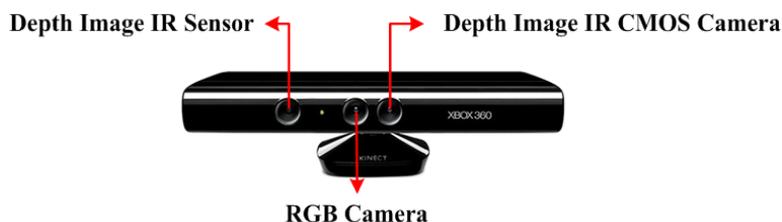


Figure 1. The appearance of the Kinect [6]

In this paper, the image depth information is used to extract the hand position, and it relates with the distance of the hand and Kinect only. It has strong robustness to the background interference and less calculation, and can meet the requirements of real-time fitting.

II. ACQUISITION THE SKELETAL POINTS

Skeletal tracking technology is through processing the depth of field data to build the human body each joint coordinate. Skeletal tracking can determine the position of various parts of the body, such as the hand, head, and the

body. Skeletal tracking is to produce X, Y, Z data to determine these skeleton points. In the 3D coordinates of joints points it abandons the Z value, only uses the X, Y value. The Kinect manages to provide the depth data of each node (Z value), but it can't use, this looks very waste. We use the node Z value, and it is not used directly, and is not shown on the interface of UI. In the coordinate space,

the depth data is needed in. The Z value is in the same scale of image, and it can be found that the size of the image is of the reverse Z. That is to say, the depth value is smaller, and the bigger images, namely the figure is closer to Kinect, the skeletal data is more.

The 20 joints traced by Kinect are as shown in Fig .2.

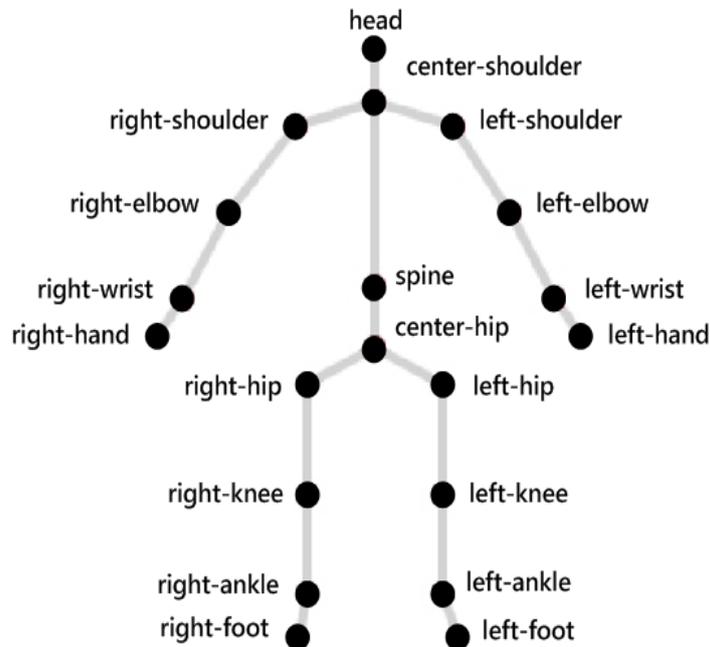


Figure 2. The skeleton points tracked by Kinect

III. ACQUIRING THE HAND POSITION

A. OpenNI

OpenNI [7-8], that is open natural interaction, includes voice, gestures and body movement. That is to say, it is hoped to be able to control some equipment, needing not other equipment but intuition only. OpenNI itself defines the API (Application Programming Interface) for writing the natural operation procedures required, and provides a multi language platform framework, and further it provides a standard interface, which can be used to obtain and analyze the data for the development.

OpenNI application can be divided into three layers. The upper layer is application, which is used to obtain the data for developing corresponding to the platform. The intermediate provides interfaces for OpenNI, and it also shows what the OpenNI can do. The bottom is the hardware part, which is including: 3D Sensor, RGB Camera, IR Camera and Audio Device. At present, as the intermediate layer of OpenNI, mainly defines four components: Full Body Analysis, Hand Point Analysis, Gesture Detection and Sense Analyzer.

In OpenNI, production node is used to represent the basic unit, and production node in production node, is for detecting and tracking the hand. When the hand area is detected, it will generate a notification message.

B. Hand Acquisition Algorithm Description

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

```

for (I = 0; mHands.getSize(); i++)
{
    if (mHand is new)
    { print message tracking; }
    else if (mHand is lost)
    { print tracking lost message; }
    else if (mHand is in tracking state)
    {
        Getting the hand coordinates;
        Translating the 3D coordinates into the hand
        position coordinates in depth coordinates;
    }
}

```

IV. POINT MATCHING ALGORITHM

A. Point Matching

Common point matching algorithm includes: diamond search algorithm [9], hexagon search algorithm [10], MVFAST [11] and PMVFAST [12].

In the system, according to the motion vector of connected block, by using the formula Fig .1 and Fig .2, it predicts the initial search point of the current block. Due to local connected point with motion vector is similar, the connected block is as the reference vector, and it uses the formula Fig .3 to dynamically adjust and judge static block threshold. By calculating movement intensity of the current connected block, it predicts the current block in large motion, or small motion. Through the above judgment, it uses the mixed templates for motion vector search.

$$\text{Point}_x = M_Val(mVect1_x + mVect2_x + mVect3_x) \quad (1)$$

$$\text{Point}_y = M_Val(mVect1_y + mVect2_y + mVect3_y) \quad (2)$$

$$\text{Thr} = \max(\text{Thr}_{\min}, \min(\text{Thr}_{\max}, \text{Thr}_{SAD})) \quad (3)$$

B. Algorithm Description

- Using Fig .1 and Fig .2 to calculate the initial search point (x, y).
- According to Fig .3, it calculates the dynamic SAD, and determines the macro block vector (x, y)

is a static block or not; if it is a static block, then jump to (6), or jump to (3).

- Calculating the direction type and judging horizontal, vertical, horizontal or non-vertical type.
- To compute the motion type: small motion, moderate motion or large motion.
- To calculate the motion vector.

V. SOMATOSENSORY FITTING SYSTEM

As shown in Fig .3, the system is divided into two parts: the hand point recognition and fitting control. In the hand point recognition part, by downloading the client and installing the Kinect driver, the depth data and RGB data streams can be accessed. When the depth data flow gets into the hand recognition module of OpenNI, three-dimensional coordinates of the hand point are obtained, and then 3D coordinates are transformed into hand position coordinates in the depth coordinates. By the point matching algorithm, left hand point coordinate matches the clothes, and if it is true, human body matches clothes in the virtual fitting room. By the point matching algorithm, right hand point coordinate matches the function button center point, and if the match is successful, the corresponding function is implemented. In the fitting control part, human body coordinates can be accessed by Kinect, when it receives the left hand successfully matching instruction, clothes is matched with human body coordinates to achieve the fitting effect, and when it receives the right hand successfully matching instruction, and the corresponding function executes in the virtual fitting room.

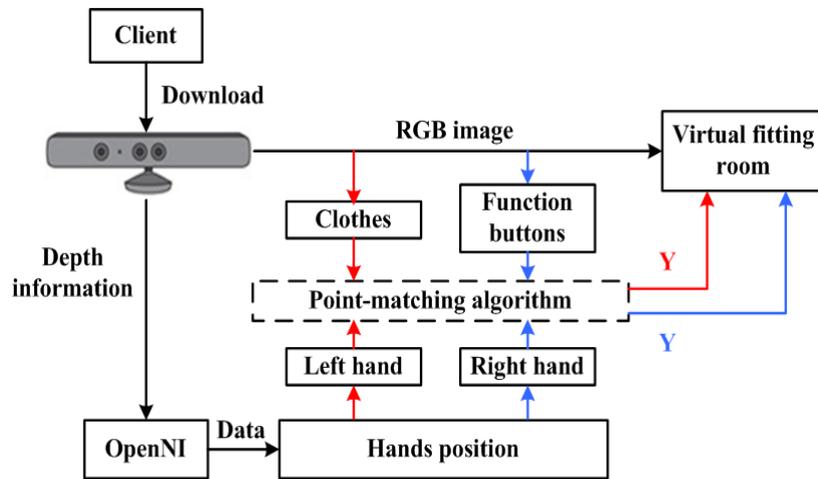


Figure 3. The design of somatosensory fitting system

VI. ANALYSIS OF EXPERIMENTAL RESULTS

A. Development Environment

Experimental environment: a Kinect, Kinect driver package ,OpenNI Kit, a virtual electronic commerce website, C++ programming language, Visual studio 2010.

B. The Experimental Process

Fig .4 is the experimental process. The interface is composed of two parts. The left part is the clothing pictures, which are from the shopping site of simulation experiment. On the right is the four function buttons. Fig .4 (a) - (a) is the cancel button that is to end the fitting. Fig .4 (a) - (b) is a sound controlling button, namely the fitting can be controlled by sound control. Fig .4 (a) - (c) is the back page button, and, according to the number of clothes

on the website, it realizes the back flip function. Fig .4 (a) - (d) is the next page button, and it realizes the page down function.

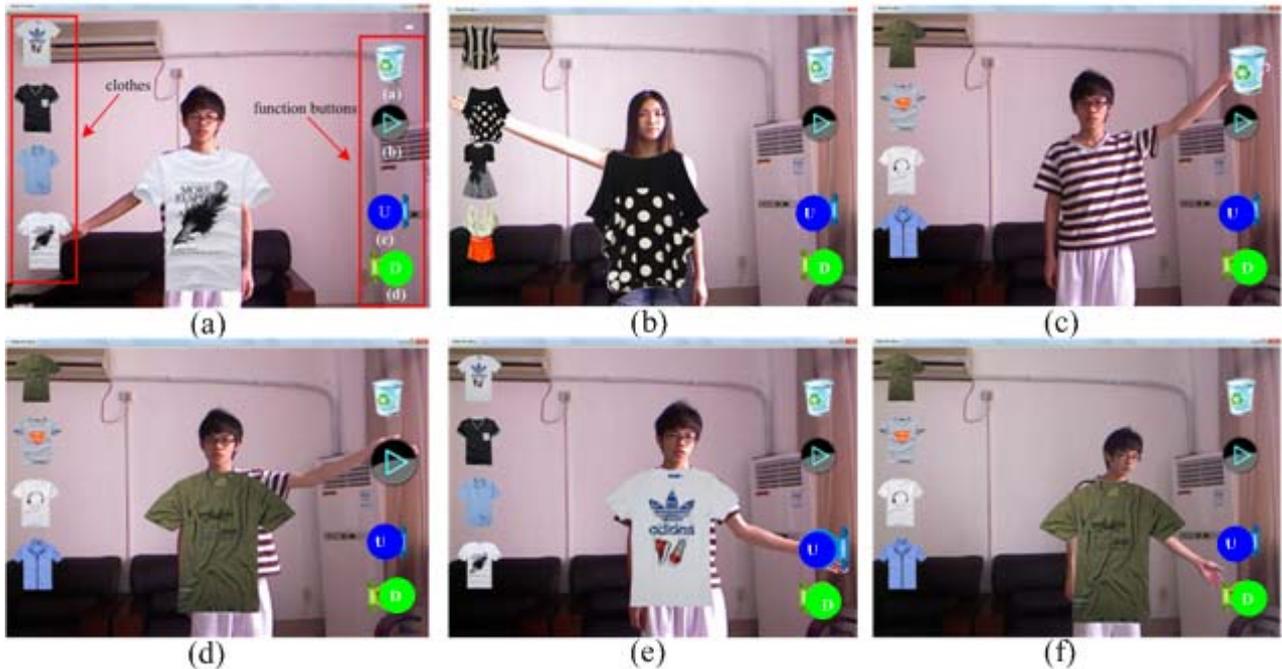


Figure 4. Experimental process

C. Experimental Procedures

- To login the simulated shopping website.
- To download the Kinect client, connect the Kinect somatosensory peripheral to PC.
- To run the system, enter the interface as shown in Fig .3.
- The user is standing in front of Kinect 1.5m-2.0m, and uses the hand to select.
- To match clothing on the left side of the working interface in Fig .3.
- Selecting functions of Fig .3 on the right side of the interface.

D. The Experimental Conclusion

The experiment was carried out in randomly selected 20 people for fitting, and the matching rate is 100%, indicating that the system has higher stability. The average time is 0.38 seconds, so the system meets the real-time requirements.

VII. CONCLUSION

In the precondition of obtaining the hand coordinates by using Kinect sensor, it uses the dynamic threshold for the static block determination, and then judges the direction of motion and motion types, and finally it is applied to the mixed template for motion estimation, and gets the matching results to control the fitting. The experimental results show that, this system not only has good stability, but also meets the real-time network fitting.

REFERENCES

- [1] Li G.S., J.H.Luo and J. Zhang.Design and realization of somatosensory dress fitting system based on Kinect [J] (In Chinese). Tsinghua Sci.Technol., , 2012,3(Supplement):48-50.
- [2] DENG Rui, ZHOU Ling-ling, Ying Ren-dong. Gesture extraction and recognition research based on Kinect depth data [J] (In Chinese). Application Research of Computers, 2013, 30(4):1263-1274.
- [3] Xia Kaijian, Wang Shitong.SKIN DEFORMATION SIMULATED WITH IMPROVED SMOOTH SKINNING ALGORITHM [J] (In Chinese). Computer Application and Software, 2009, 26(12):174-176.
- [4] LI Hongbo, DING Linjian, RAN Guangyong.Analysis of human identification based on Kinect depth image [J] (In Chinese).DIGITAL COMMUNICATION, 2012, 4:21-26.
- [5] ZHANG Di.Discussion of the Kinect Applications [J] (In Chinese).LOGISTICS ENGINEERING AND MANAGEMENT, 2012, 34(6):39-41.
- [6] ZHANG Yi, ZHANG Shuo, LUO Yuan, XU Xiao-dong. Gesture track recognition based on Kinect depth image information and its applications [J] (In Chinese). Application Research of Computers, 2012, 29(9):3547-3550.
- [7] [EB/OL]. <http://www.openni.org/>.
- [8] [EB/OL]. http://www.openni.org/software/?cat_slug=file-cat1.
- [9] [9]Zhu S, Ma K K. A new diamond search algorithm for fast block matching motion estimation [J], IEEE Transactions on Image Processing, 2000, 9(2): 287-290.
- [10] Zhu C, Lin X, Chau L P. Hexagon-based search pattern for fast block motion estimation [J]. IEEE Transactions on Circuits and System for Video Technology, 2002, 12(5) : 349-355.
- [11] Hosur P I, Ma K K. Motion vector field adaptive fast motion estimation [A]. In: Proceedings of Second International Conference on Information, Communications and Signal Processing [C], Singapore, 1999: 7-10.
- [12] Tourapis A M, Au O C, Liou M L. Fast Block-matching Motion Estimation Using Predictive Motion Vector Field Adaptive Search Technique (PMVFAST). [EB/OL] <http://users.softlab.ece.ntua.gr/~tourapis/navigatate.htm> ,2000-03.