Real-time Measurement of Pointing Action by Using DSP

Takeshi Shiofuku , Norihiro Abe , Yoshihiro Tabuchi
Kyushu Institute of Technology
680-4 Kawadu, Iizuka, Fukuoka 820-8502, Japan
Tel : 81-948-29-7776
shiofuku@sein.mse.kyutech.ac.jp

Hirokazu Taki
Wakayama University
930 Sakaedani, Wakayama-shi
Wakayama 680-8510, Japan
taki@sys.wakayama-u.ac.jp

Shoujie He
Eastman Kodak Company,
Plano, Texas, USA
hesj@computer.org

Abstract:
This paper presents an approach of measuring in real-time the vector of finger that is pointing to an object. DSP is used in the operation processing unit in order to do the real-time processing. The steps include the extraction of flesh-colored regions from an image, the labeling of the flesh-colored regions, and the detection of two characteristic positions on the finger so that the direction that the finger is pointing at will be calculated. The entire process takes about 29msec, which makes it possible to have the frame rate of 34fps. With this frame rate, this measurement approach is considered real-time and promising to be merged into other application systems.

Keywords: **DSP, Real time, Pointing action,**

1. INTRODUCTION

With the advancement of data processing and communication technologies, more and more people are using the Internet in their daily life. So far, the Internet was mainly used for collecting information and sending or receiving email messages. With the widely deployed high-speed communication networks that are using fiber optics technologies and the lower price but high-performance computers, it becomes possible to telecommunicate the data-intensive content such as images and voice through the Internet.

Video conference is an application that involves the telecommunication of voice and video images. People attending the conference can communicate with each other even if they are physically located in a separate place. Such a video conferencing system is very convenient and widely used tool. However, voice and video images are sometimes insufficient when people are trying to express their ideas with gestures such a pointing action. It is considered more preferred if the computer could help with the recognition of a point action.

The interactive shopping support system with avatar in a virtual space has been studied in the past [3][5]. Users can interact with the system with voice and hand motion. The system is made easy to use, especially for the convenience to the senior citizen and children who are unfamiliar with the regular interfaces of a computer system. In this system, nonverbal processing is required for the recognition of voice and hand motion.

This research aims at developing an approach for the measurement of the 3D direction of a pointing action in real-time. DSP is used for achieving high performance in terms of the speed.

2. System Configuration

A DSP board and two CCD cameras are used. The application program is written in Delphi. The maximum frame rate of CCD camera is 60fps.

2.1 DSP Board

The DSP board is a board with semiconductor chip that can processes the digitalized signal at a high speed. The operation processing part of the DSP board has 4 DSP which can process at a speed of 5BOPS. One BOPS means one billion times operations per second.

2.2 Application Communication

The data and command communication between a PC and the DSP board is achieved through an application program from Delphi. To use the DSP, the program written in Delphi needs to load the library of execution files.

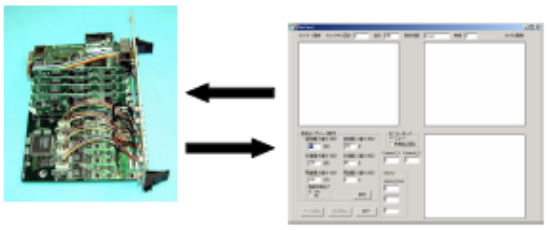

Fig1: Communication of Delphi application and DSP

2.3 Research Environment

Figure 2 shows the research environment. The two cameras are set at the right side with the optical axes in parallel. No flesh-colored object is supposed to be put in the background. The distance between the display and the reference point of finger is 400mm.

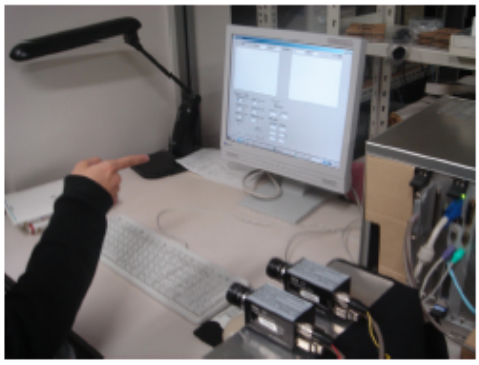

Fig2: Research Environment

3 Pointing information acquisition

In order to detect the pointing direction, the system must extract flesh-colored regions from the image and assign a label to each region by removing noises. To calculate the vector of a forefinger, two characteristic points on the forefinger are detected.

3.1 HSI conversion

The original image data obtained from the camera are in RGB color system. It is, however, difficult to set the range of the target colors in this system.

This problem could be resolved by using the HSI color system. In the HSI system, the person can express the color sensuously. It is necessary to convert the image data from the RGB color system to the HSI color system. In the HSI color system, it is empirically known that a flesh-colored area is within the following ranges. Figure 3 shows the result with the flesh-colored region extracted.

Hue : $-75 < H < 40$ (Range : 0~360)
Saturation : $40 < S < 200$ (Range : 0~255)
Intensity : $0 < I < 255$ (Range : 0~255)

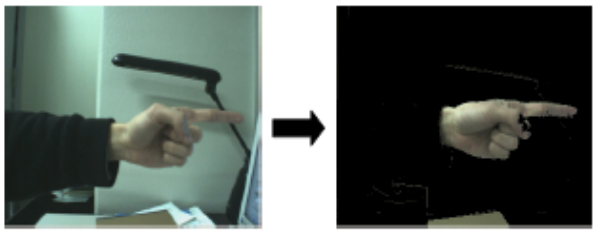

Fig3: The extracted flesh-colored image

3.2 Labeling

Labeling is a process of assigning labels to each of the connected components. In this research, we used the run-length method with which the regions with a small area were removed.

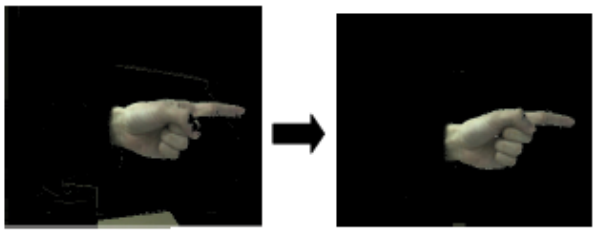

Fig4: The labeling

3.3 Characteristic Point Detection

The first characteristic point of a finger is the rightmost point. The other point is decided by comparing the amounts of the pixel forming the finger tip to back the hand.

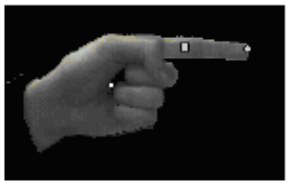

Fig5: Characteristic Point

The comparison equation is as follows. The index i is taken in a horizontal axis and x_dis[i] shows amount of pixels.

$$x\_dis[i] > x\_dis[i+5] * 1.5 \qquad (1)$$

3.4 Stereo vision

Stereovision is a method of observing the same object from two different view points and measuring three dimension position of the object by the parallax of characteristic points.

As shown in Figure 6, the three-dimensional coordinates of object P can be measured with the formulae (2), (3) and (4). Length in Figure 6 shows the distance between cameras, and f shows the focal length of the cameras.

$$X = Length\frac{(XL + XR)/2}{XL - XR} \quad (2)$$

$$Y = Length\frac{YL/2}{XL - XR} \quad (3)$$

$$Z = Length\frac{f}{XL - XR} \quad (4)$$

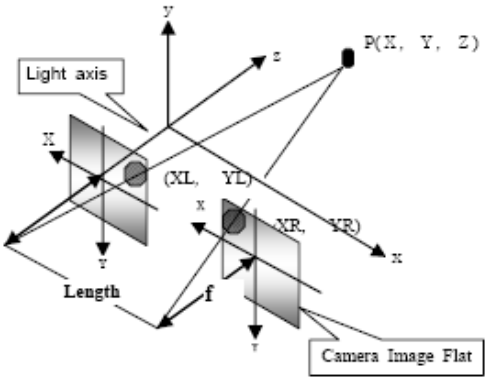

Fig6: Three-dimensional positional coordinates

3.5 Flow Chart

Figure 7 shows the flow chart with the above-mentioned processing.

4. Experiments

4.1 Experiment Methodology

1: Actually, even if a forefinger is still, the designated point cannot be fixed at one point. Therefore, a range in which the point moves around will be detected.

2: The processing time for the detection of characteristic points was measured.

4.2 Experimental Results

1: Figure 8 shows target points on a plane that is placed at the location of 45 cm from a forefinger during the measurement of a pointing action. The coordinates of the median point of this dataset is (160,120). This means that the target position is contained in a region of 25 x 13pixels. Not only numeric data but also the collection of the points is displayed as shown in Figure 9.

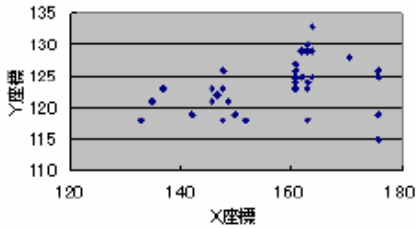

Fig8: variance of a target position

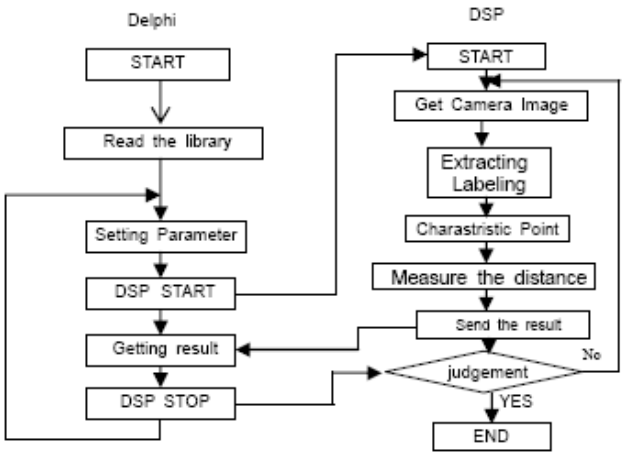

Fig7: Flow Chart

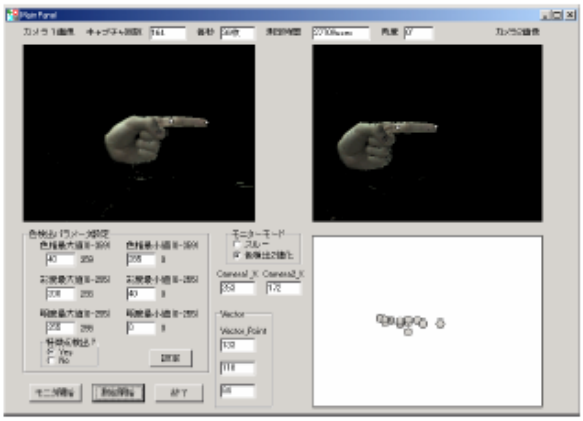

Fig9: Application

2: Time taken to capture an image is about 29msec and number of frames per second is about 34. Figure 10 shows the results of the measurement.

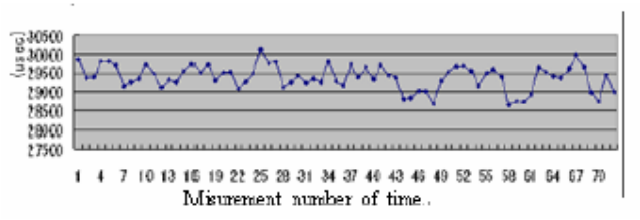

Fig10: The processing time

6. Future Work

The problem with this approach is that the computer failed to extract hand and only hand when there is a flesh-colored object behind the hand. One of the future work will be to make the hand extraction more robust and accurate. A possible approach could be using correlational method or calculus of finite differences.

7. Conclusions

Time needed to capture an image is less than 30msec. A target point is within 25pixels from the median point. Since this method helps a user to specify an object on a plane placed at 45cm from the tip of his forefinger, it is possible to user a forefinger as a pointing device. Using this method together with a voice allows user to specify an object without spelling out its attributes such as name or color.

8. Acknowledgement

We greatly appreciate the aid of Ministry of Internal Affairs and Communications (MIC) and the Grant-in-Aid for Scientific Research.

9. REFERENCES

[1]Yoshiaki Shirai , Masahiko Taniuchida
"Pattern information processing"

[2]Yasuhiro Watanabe, Norihiro Abe, Kazuaki Tanaka, Hirokazu Taki, Tetsuya Yagi,
Multimodal communication system allowing man and avatar to use voice and beck,
3rd International Conference on Information Technology and Applications (ICITA'2005), pp.161-166

[3] Tatsuya Yoshikawa, Syunji Uchino, Norihiro Abe, Kazuaki Tanaka, Hirokazu Taki, Tetsuya Yagi, Shoujie He:
Voice and gesture recognition system facilitating Communication between man and virtual agent, INVITE2006 International Conference on Parallel and Distributed Systems), Vol.2, pp.673-677, 2006

[4] Tetsuya Sato, Norihiro Abe, Kazuaki Tanaka, Yoshimasa Kinoshita, and Shoujie He:
"Toward Developing Multiple Organs and Diseases Diagnosing Intellectual System referring to Knowledge Base and CT Images", accepted, CBMS 2006(19th IEEE International Symposium on Computer-Based Medical Systems),
pp.359-364,2006

[5]Syunji Uchino, Norihiro Abe, Shoujie He, Hirokazu Taki, "Real-time Interactive Dialog System between Person and Virtual Agent", accepted, International Symposium on Artificial Life and Robotics (AROB 2007)