A Three-Dimensional Fingertip Interface

Yangkeun Ahn[1,2], Kwangmo Jung[2], and Jiman Hong[1]

[1] Soongsil University, Seoul, South Korea
[2] Korea Electronics Technology Institute, Seoul, South Korea
ykahn@keti.re.kr

Abstract. This paper proposes a method to recognize the fingertip three-dimensionally using infrared stereo cameras. The proposed method can be used for human-computer interactions with a three-dimensional display and is designed to provide robust performance against finger trembling, kinematic errors, and sensor noise. This paper describes the proposed method in detail and also presents implementation results.

Keywords: Fingertip Recognition, Stereo Camera, Spatial Touch.

1 Introduction

This paper proposes a three-dimensional interface using a fingertip with no separate equipment to ensure interactions between the three-dimensional image projected onto the space and the user. The image projection system discussed here is a device that projects a two-dimensional image onto the concave mirror, resulting in binocular parallax and enabling the user to gain a sense of visual depth. This system provides the user with an immersive interface. The use of the existing touch screen technique to ensure interaction between the visualized image and the user entails the following problems: First, the input device is exposed to the user when a physical touch screen is installed, thus undermining the user's sense of immersion. Second, the touch screen equipment should recognize the depth of the finger, as the image perceived by the user is not two-dimensional, but conventional touch screens are incapable of recognizing depth.

In order to provide the user with an immersive interface, therefore, an optical method should be chosen such that the existence of the touch screen is unknown to the user. Also, the location of the user's finger should be identified in a three-dimensional space so as to facilitate interaction with the user. This paper proposes a method using two infrared cameras to identify the location of the finger in three-dimensional space (see Fig. 1). In order to implement the proposed method, this paper discusses detection of the location of the fingertip within images, alignment of the fingertip points of two images to calculate the three-dimensional location, and the design of a linear Kalman filter to deal with image noises.

2 Methodology and Test Details

The method proposed in this paper can be summarized as follows: First, infrared light is projected onto the finger using the infrared emitter. Second, the images of the

C. Stephanidis (Ed.): Posters, Part II, HCII 2011, CCIS 174, pp. 141–145, 2011.

reflected finger are recorded with infrared stereo cameras. Lastly, the pixel coordinates of the finger extracted from the two images and the cameras' intrinsic parameters are used to calculate the three-dimensional location of the finger. The overall process is described in Fig. 2. This section introduces details of each method.

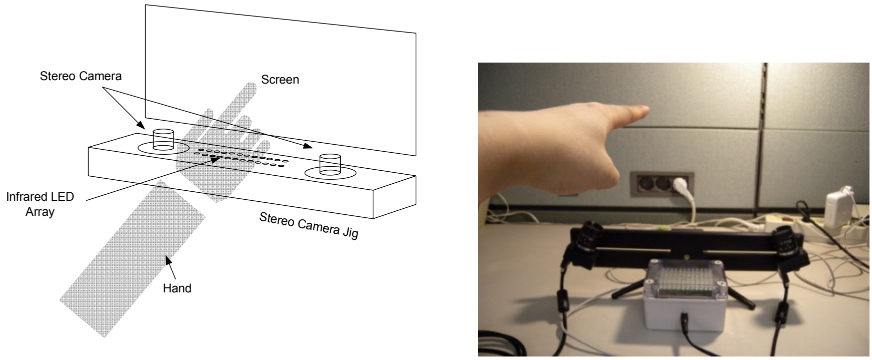

Fig. 1. Conceptual diagram of proposed device and configuration of testing equipment

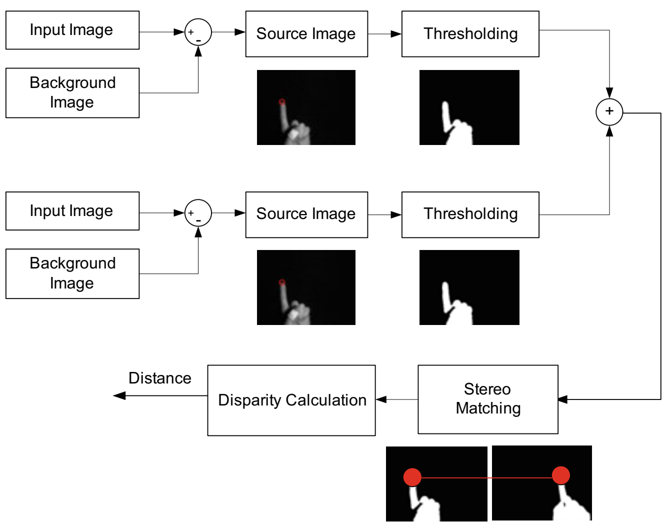

Fig. 2. Conceptual diagram of processing process

2.1 Configuration of Testing Equipment

The testing equipment used in this paper is shown in Fig. 1. The infrared stereo cameras are used here to minimize errors caused by the subject(s) other than the finger. The impact of noise can be reduced, as infrared wavelengths alone are accepted from near distance. Two Chameleon (a model from Point Grey) cameras are installed on the stereo jig, as shown in Fig. 1. An infrared radiation device is used to project infrared light onto the finger and extract the finger's image.

2.2 Extraction of Fingertip Image

This section explains how to extract the image of the fingertip from the fed images. This method consists of hand area extraction, binarization, fingertip area extraction, and Kalman filter-based stabilization. When the system operates, the background image of the environment, where the equipment is located, is obtained by showing the hand in the image. The obtained background image is used in the difference operation with a subsequently entered image to remove the background from the latter image, and the resulting image consequently contains the hand area reflected by infrared light only. The extracted image of the hand is a gray-scale image with each pixel having a value from 0 through 255. Binarization is carried out to clearly distinguish the finger area from the remaining area.

2.3 Extraction of Fingertip Area

To identify the coordinates of the fingertip from the binarized hand area, a filter bank is created based on human observations. The filter bank is a method to connect filters with different conditions in series and obtain desired results in each phase to ultimately gain the fingertip's image. This method breaks the geometrical forms of the fingertip into three phases in total, producing powerful results from a combination of simple rules. The geometrical characteristics of the fingertip can be summarized as follows [1]:

- The finger area remains white within the binarized image;
- The whole of the small circle S is included in the white area in Fig. 3; and
- The large circle L cuts the fingertip area in two points only (i.e., points a and b) in Fig. 3.

The method used here applies the aforementioned three filters consecutively to take the ultimately remaining area as the fingertip area candidate. The algorithm for this method can be expressed as follows:

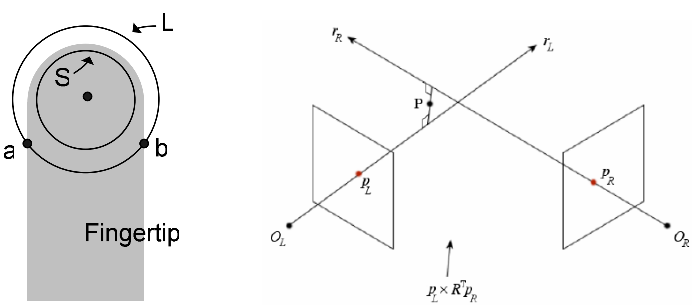

Fig. 3. Geometric analysis of fingertip and optic axes of two non-transversal cameras

```
For each pixel (x,y):
    For each filter:
        If i-th rule at pixel (x,y) is false
            Skip pixel;
        Else
            Mark (x,y) as fingertip
        End if;
    End for;
End for;
    Return fingertip candidates;
```

2.4 Calculation of Three-Dimensional Location of Fingertip

This paper assumes that the cameras are calibrated in advance. The intrinsic parameter matrices of the two cameras, obtained through calibration, are denoted as $\mathbf{K}_l, \mathbf{K}_r$ here; all locations are specified with the left camera as the basis.

Fig. 3 describes how the optic axes of the two cameras do not exactly intersect with the fingertip points extracted from the stereo vision. Thus, the minimum distance crossing the two rays orthogonally should be derived [2].

The extrinsic parameters of the calibrated stereo cameras are referred to as R_l, R_r, T_l, T_r, which represent the left rotating matrix (=I), the right camera's rotating matrix in relation to the left side, the location of the left camera (=[0,0,0]), and the right camera's location with the left camera as the basis, respectively. The new parameters of the stereo vision can be written as below:

$$\begin{aligned} R &= R_r R_l^T \\ T &= T_l - R^T T_r \end{aligned} \quad (1)$$

To identify the three-dimensional location of the finger using new parameters, R, T, the following linear system should be solved to derive the parameters a, b, c:

$$a\mathbf{p}_l - bR^T\mathbf{p}_r + c(\mathbf{p}_l \times R^T\mathbf{p}_r) = \mathbf{T} \quad (2)$$

To solve the above equation, the matrix $\mathbf{A}$ is derived as given below to set up the linear system:

$$\underbrace{\begin{bmatrix} \mathbf{p}_l & -R^T\mathbf{p}_r & \mathbf{p}_l \times R^T\mathbf{p}_r \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \mathbf{T} \quad (3)$$

The parameter set, $\mathbf{x} = [a, b, c]^T$, can be derived from $\mathbf{A}^{-1}\mathbf{T}$, and the location of the fingertip can be determined on this basis using $a\mathbf{p}_l$.

2.5 Test Results

The system is used to test the user's inputs. When the user enters rectangular, triangular, and circular movements, the results also resemble those movements. Performing computation 20 times per second, the system can be reliably used as an interface.

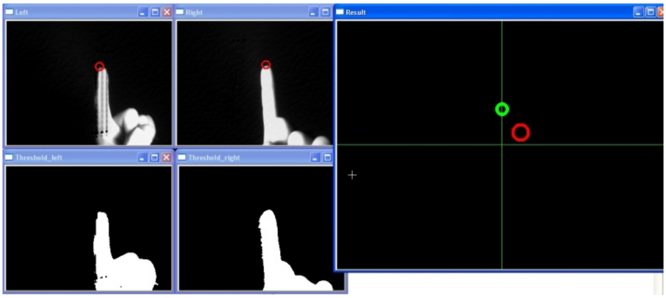

Fig. 4. Test results

3 Conclusion

This paper has proposed an infrared stereo camera-based method for three-dimensional fingertip recognition. Using the proposed method, we identified the location of the fingertip with the series filter bank, stabilized it using the Kalman filter, and obtained the three-dimensional coordinates of the fingertip with the left camera as the basis.

To assess the level of accuracy in the future, it will be essential to videotape already known movements and compare the two results. Also, research is needed on a correction algorithm designed for 1:1 correspondence with the designated image.

References

1. Letessier, J., Berard, F.: Visual Tracking of Bare Fingers for Interactive Surfaces. In: UIST 2004, ACM, New York (2004)
2. Trucco, E., Verry, A.: Introductory Techniques for 3-D Computer Vision. Prentice Hall, Englewood Cliffs (1998-2006)