Vision-based Input Interface for Mobile Devices with High-speed Fingertip Tracking

Yuki Hirobe   Takehiro Niikura
Yoshihiro Watanabe   Takashi Komuro   Masatoshi Ishikawa
Graduate School of Information Science and Technology, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8638 Japan
{Yuki_Hirobe, Takehiro_Niikura}@ipc.i.u-tokyo.ac.jp
{Yoshihiro_Watanabe, Takashi_Komuro, Masatoshi_Ishikawa}@ipc.i.u-tokyo.ac.jp

ABSTRACT
We propose vision-based interface for mobile devices by tracking 3D motion of a human fingertip using a single camera and by recognizing small input gestures in the air. Since the fingertip near the camera moves fast in the image, a high-frame-rate camera is used for stable tracking. In order to separate the click gesture from the others, frequency filter is applied to scale change in the fingertip images.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. - Input devices and strategies.

General terms: Algorithms, Design

Keywords: Vision-based UI, portable device, 3D input interface

INTRODUCTION

Recently mobile devices have become small and are difficult to have input interface that has wide operation area on their surface. For example, conventional interface such as a keypad or a touch panel on a cell phone has limited operation area.

There has been research on input interface systems that overcome this problem, but they require users to wear some physical devices [1] or require flat surface for operation [2].

Therefore we propose vision-based 3D input interface for mobile devices which does not require wide space on the device's surface, other physical devices or specific environments. Users can operate the device by the movement of a fingertip in the air.

There are some technological issues in order to realize such an interface system.

- Separation of fingertip region from an image with varying background
- Stable recognition of the position and posture of a fingertip which moves fast
- Accurate detection of click actions using a single camera

UIST'09, October 4–7, 2009, Victoria, BC, Canada.

In this research, the system binarizes an image of the fingertip and uses it as a template image. Then it tracks the fingertip in input images about translation, rotation and scale. In that way, the positions and postures of the fingertip are estimated. In addition, a high-frame-rate camera enables stable tracking of the fingertip that moves fast in the image. Small click gesture in the air is detected by applying a frequency filter to the scale change of the fingertip images in order to pick up its distinctive movement.

SYSTEM COFIGURATION AND OPERATION

System Configuration

We constructed a system shown in Figure 1. A compact IEEE1394 high-frame-rate camera Firefly MV (Point Grey Research Inc.) with a lens having a focal length of 1.9 mm, PC and a small USB display. We used the camera at a frame rate of 74 fps (frame per second) with an image size of 400 x 160 pixels.

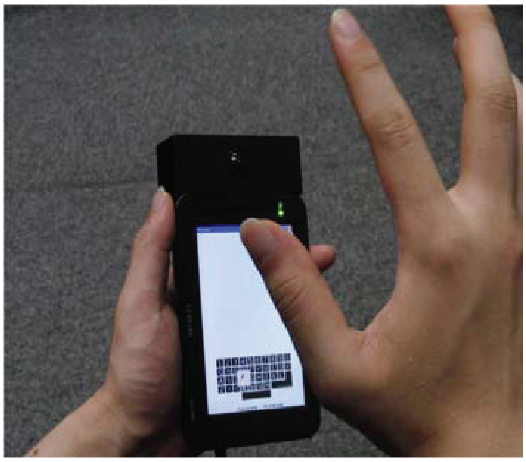

Figure 1: Developed System

Fingertip Tracking

A fingertip moves in three-dimensional space and the shape of the fingertip in the image is transformed in various ways. In this interface, transformation of the binarized fingertip region image mainly consists of translation along the plane perpendicular to the camera's optical axis, rotation around the optical axis and scale change which is related with the finger's distance from the camera.

Therefore transformation of the image can be represented by 4 parameters $\mathbf{p} = (t_x; t_y; s; \theta)^T$. t_x and t_y are translation amounts on an image plane, θ is a rotation amount around the optical axis, s is a scale amount. $\mathbf{x} = (x; y)^T$ indicates an image coordinate. The transformation of the image is represented as the following equation .

$$\mathbf{W}(\mathbf{x};\mathbf{p}) = \begin{pmatrix} s\cos\theta & -s\sin\theta \\ s\sin\theta & s\cos\theta \end{pmatrix} \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} + \begin{pmatrix} sx_0 + t_x \\ sy_0 + t_y \end{pmatrix} \quad (1)$$

x_0,y_0 are the coordinates of the center of rotation.

The Lucas-Kanade Algorithm [3] is used to estimate these parameters.

As shown in Figure 2, the template image for tracking is taken from the binarized image of the user's fingertip inside the box in the input image when tracking starts.

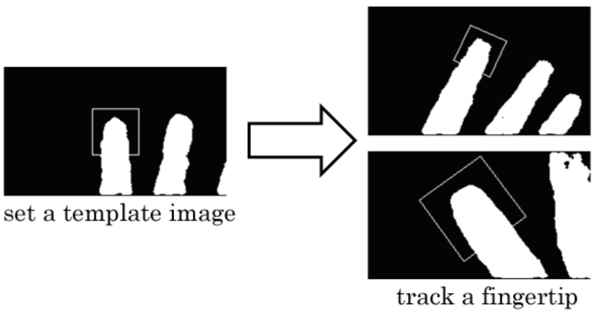

Figure 2: Template-matching-based Tracking

Detection of the Click Action

It is desirable that input actions should be instinctive. We defined the following gesture as input actions. The gesture is slightly moving a fingertip back and forth like clicking a mouse's button in the direction of the camera's optical axis.

At that time, the click action needs to be distinguished from the other fingertip movement. Therefore utilizing the difference of the dominant frequency of the scale, the click action can be detected by using the scale transition with a frequency filter.

APPLICATION EXAMPLES

As described above, the interface we propose can estimate 3D position of the fingertip and detect click action. This characteristic enables various operations on mobile devices. We constructed three application examples.

In-air Typing Keyboard

In this application, the user can input text with software keyboard, as shown in Figure 3(a). When the pointer, which moves according to the position of the fingertip, is located on the target key and click action is carried out, the target character is input. Click actions can be detected by using the scale transition with a frequency filter.

Zooming Picture Viewer

The user can zoom and scroll the picture on the display with the 3D position of the fingertip in this application, as shown in Figure 3(b).

The user can grab the picture by one click action, and then the picture zooms and scrolls on the screen according to the 3D position of the fingertip. When a click action is carried out again, the picture is released and fixed on the screen.

3D Painter

This interface can estimate the 3D position of the fingertip accurately with a single camera. In this application, the position is directly used. As shown in Figure 3(c), the user can draw lines in 3D space with fingertip. The lines are drawn according to the fingertip 3D position.

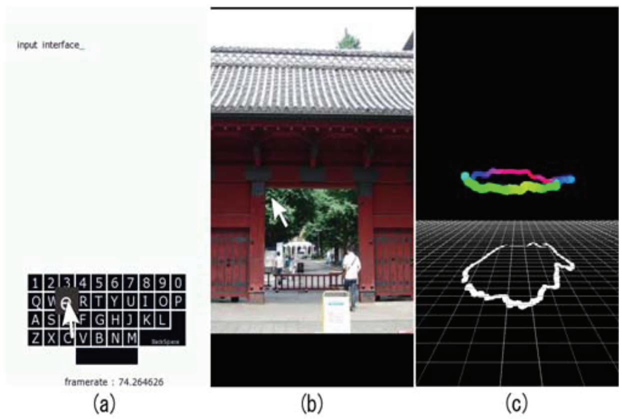

Figure 3: Application: (a) In-air Typing Keyboard (b) Zooming Picture Viewer (c) 3D painter

CONCLUSION

In this paper, we proposed vision-based 3D input interface for mobile devices which allows a user to input with wide operation area by tracking the 3D movement of the fingertip and by detecting small click actions in the air. We showed three applications, In-air Typing Keyboard, Zooming Picture Viewer and 3D painter.

REFERENCES

1. Kölsh, M. and Turk, M. Keyboards without Keyboards: A Survey of Virtual Keyboards. In *Proceedings of Sensing and Input for Media-centric Systems,* 2002.
2. Roeber, H., Bacus, J. and Tomasi, C. Typing in Thin Air The Canesta Projection Keyboard - A New Method of Interaction with Electronic Devices,*CHI '03 extended abstracts on Human factors in computing systems*, 2003, pp. 712-713.
3. Baker, S. and Matthews, I. Lucas-Kanade 20 Years On: A Unifying Framework, *International Journal of Computer Vision*, 2004, vol.56, No3, pp. 221-255.