MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11 - 13, 2002, Nara- ken New Public Hall, Nara, Japan

3—15

Wearable Virtual Tablet: Fingertip Drawing Interface using an Active-Infrared Camera

Norimichi Ukita, Akihiro Terabe, Yasuyuki Kono, and Masatsugu Kidode
Graduate School of Information Science, Nara Institute Science and Technology, Japan
{ukita,kono,kidode}@is.aist-nara.ac.jp

Abstract

We always input various kinds of information into a computer when using it. As wearable technology advances, several works about portable interfaces have been reported. These interfaces, however, have some disadvantages: 1) it is impossible to input free contents, for example, not only characters but also arbitrary pictures and symbols, or 2) for free-content input, a user has to use a special pen device and learn how to use it. In this paper, we propose the *Wearable Virtual Tablet* (WVT), where a user can draw a locus on an arbitrary plane object (e.g., a notebook and a magazine) with the fingertip. The WVT has the following advantages: 1) draw the locus by the fingertip without any mechanical devices, and 2) switch the input state from on/off to off/on without any mechanical switches.

1 Introduction

In this paper, we present a wearable vision system that enables us to input information into a computer by fingertip drawing. Although the following free-content input systems have been also developed, they have some disadvantages for a wearable convenient interface:

Pen device with physical sensors[1]: A user possesses a pen device with gyro and acceleration sensors. By employing these sensors, the 2D locus of the pen device in the air can be tracked. This system, however, has the following disadvantages: (1) the user has to always possess the pen device, and (2) the user has to turn on/off a mechanical button for switching the input state from on/off to off/on.

Fingertip writing with a single stroke[2]: A user writes a character with a single stroke by moving the fingertip in the air. The finger motion is observed by the head-mounted camera, and the 2D fingertip locus is displayed on the HMD. This system, however, has an essential problem; the user has to write a character with a single stroke.

To solve the problems in these systems, the system should provide the following functions simultaneously:

- Draw the locus by the fingertip.
- Switch the input state from on/off to off/on without any mechanical switches.

To realize these functions, we propose the *Wearable Virtual Tablet* (WVT, in short), where a user can draw a locus on a plane object with the fingertip.

2 Wearable Virtual Tablet and Its Architecture

2.1 Basic Scheme for the WVT

The WVT system consists of a wearable camera with a HMD as shown in Fig.1. In the WVT system,

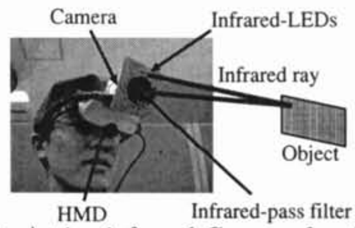


Figure 1: Active-infrared Camera for the WVT.

(a)Infrared image (b)Color image (c)Close region

Figure 2: Close-object extraction using the active-infrared camera.

we can draw an arbitrary locus on an arbitrary plane-object even while freely walking and moving our body. The position of the fingertip on the input plane is detected and tracked in the observed image. Next, its locus is superimposed on the current observed image. The fingertip locus drawn by the user is superimposed on the current observed image. The superimposed image is shown in the HMD, allowing the user to draw the locus continuously. The superimposed image is shown in the HMD, allowing the user to draw the locus continuously.

The following technical problems have to be solved in order to realize the WVT:

Problem 1 Determine the input plane.

Problem 2 Detect and track the user's fingertip.

Problem 3 Discriminate between the input and non-input states.

Problem 4 Superimpose the locus of the fingertip on the current observed image.

In this paper, to make problem 1 easy, an arbitrary rectangular plane held by the user's hand is regarded as an input plane. The system has to detect its four corners to determine the area in the image observed by the wearable camera. We use an active-infrared camera introduced in [3] as the wearable camera. Its properties drastically simplify the detection of the input plane as well as the detection of the fingertip (will be mentioned later). Furthermore, the gray-scale information in the infrared image allows the system to estimate whether or not the fingertip touches the input plane. Finally, problem 4 can be realized by translating the fingertip locus from the square buffer image for recording the locus to the input plane in the current image; based on the geometric configuration among four corners of the input plane, the projection between two tetragons (i.e., the current input plane and the buffer image) can be computed.

2.2 Active-infrared Camera

Our active-infrared camera is realized with a conventional CCD camera and infrared-pass filter; this camera can capture only infrared rays as a gray-scale image as shown in Fig.2 (a). Since infrared rays are irradiated from the light source near the device, the camera mainly obtains the infrared rays irradiated from the mounted light source and reflected from an object in the scene. The obtained reflected rays are captured by the camera as a gray-scale image.

In general, it is difficult to extract regions corresponding to objects in proximity to the camera[1] from a color image. In the example shown in Fig.2 (b), many background objects are observed in the captured image and make close-object extraction difficult. On the other hand, Figure 2 (a) shows the observed infrared image. The intensity of the reflected infrared rays depends on the distance from the light source to objects in the scene. By employing this property, close-object regions can be easily extracted from the observed image (Fig.2 (c)) without a complicated method for 3D depth reconstruction.

In addition to the above advantage, the depth-dependent information included in the infrared image allows us to estimate the geometric configuration of the observed objects (e.g., input plane and user's fingertip).

With these properties, we solve the four problems described in Sec.2.1 to implement the WVT.

3 Fingertip Drawing using the Active-infrared Camera

Figure 3 shows the flowchart of the WVT. The WVT has two functional modes:

Registration mode: To determine whether or not a user is drawing the locus on the input plane, a distribution of gray values of the pixels around the fingertip is registered in advance. For this registration, the problems 1 and 2 described in Sec.2.1 have to be solved.

Input mode: After the registration mode, a user starts drawing. In this mode, the problems 1, 2, 3 and 4 have to be solved.

3.1 Detection and Tracking of the Virtual Input Plane

In our system, four vertices of an input plane have to be detected to determine the area of the input plane and to superimpose the drawn locus in the observed image. Suppose the close-objects are only the input plane and the user's hand.

As mentioned before, the system can easily extract regions corresponding to close-objects (i.e., the input plane and the user's hand) in the infrared image (shown in Fig.4 (a)) without interference from a complicated background scene. Edge detection is applied to the close-object image. The result is shown in Fig.4 (b). To detect the four sides of the input plane, we employ the Hough transform[4]. The detected result is shown in Fig.4 (c). The four lines are, then, determined. Lastly, the intersection points of these lines are considered to be the four vertices (Fig.4 (d)).

[1] Hereafter, we call an object in front of and near the camera a *close-object*.

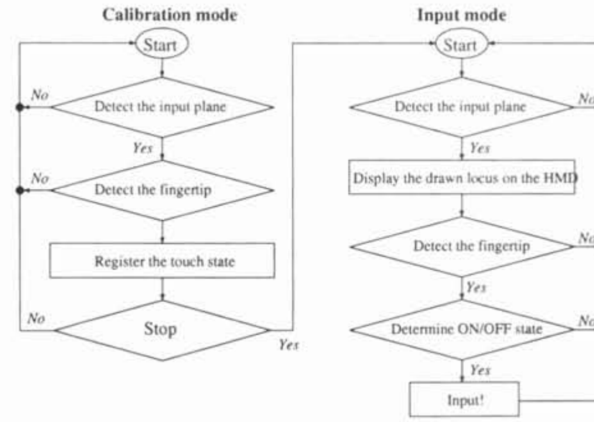


Figure 3: Flowchart of the WVT.

(a) Input image

(b) Binarized edge image

(c) Detected lines

(d) Four sides and vertices of the input plane

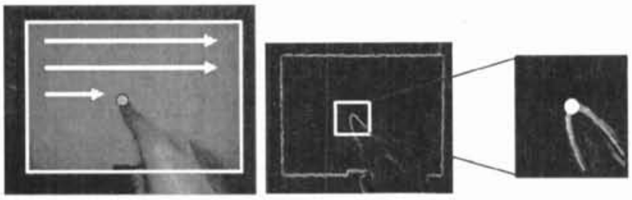
Figure 4: Detecting the vertices of the input plane.

3.2 Detection and Tracking of the Fingertip

The system searches for a fingertip position in the binarized edge image. To simplify the task, we assume that a user points his/her fingertip upward in the observed image while drawing as shown in Fig.4 and 5. To accurately estimate the fingertip position, the system computes a curvature of connected edge pixels in the input plane region as illustrated in Fig.5. The pixel corresponding to the largest curvature is considered to be the fingertip position.

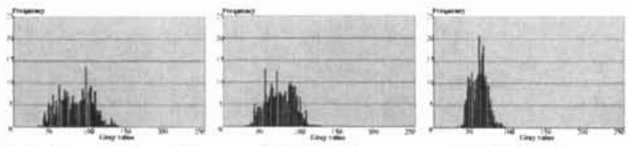
3.3 Discrimination between Input and Non-input States

Although gray values in the infrared image include the depth-dependent information, they vary not only depending on the distance but also depending on various other factors, for example, the material and posture of the input plane object. This results in difficulty in estimating the relationship between the distance and gray values for every situation in advance. Accordingly, in our system, a user registers the difference between observed gray values in the case of touch and non-touch immediately before he/she uses the WVT system with



(a) Finger search (b) Accurate position of the fingertip

Figure 5: Fingertip detection.



(a) 0[cm] (touch) (b) 3[cm] (c) 5[cm]

Figure 6: Histograms of the gray values around the fingertip.

a certain plane object.

To register the variation of gray values around the fingertip in the case of touch, the system observes the input plane while the user is drawing on it. During this operation, the system detects the fingertip region and acquires the histogram of gray values in the gradient image.

Figure 6 shows an example of the acquired histogram of gray values around the fingertip region. For comparison, we also show examples of the gray-value histograms in the case of non-touch (Fig.6 (b) and (c)), both of which were observed in the same condition, i.e., 1) the same position and posture of the input plane and 2) the same position of the fingertip in the observed image. The distributions of gray values are different from each other. We characterize the distribution as a variance of gray values. In these experimental results, as the distance between the fingertip and the input plane becomes closer, the variance of gray values becomes greater. In this case, therefore, the system detects the input state if the variance of gray values is larger than the predefined threshold. The smallest variance acquired during the above operation is considered to be the threshold.

Note that gray values in the input plane vary depending on its position and posture. This means that a constant threshold fails the detection of the input state if the input plane moves. To solve this problem, we adjust the threshold depending on the average of gray values around the fingertip. The relationship between the threshold and the gray values around the fingertip is shown in Fig.7. The horizontal and vertical axes of the graph show the average of the gray values and the variance in the edge image, respectively. This graph was obtained by observing the fingertip on the input plane while the position of the fingertip and the geometric configuration between the camera and the input plane were changed. Suppose that the smallest variances at each horizontal coordinate (i.e., the lower boundary of the observed values) are represented by the function $L(g)$ as illustrated in Fig.7. When the system is in the input mode, the system considers the fingertip as touching the input plane if the observed gray value around the fingertip is above $L(g)$.

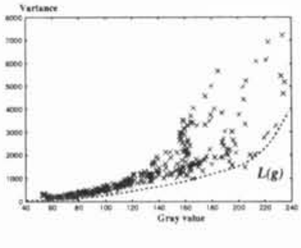


Figure 7: Adaptive threshold $L(g)$ estimated from observed values.

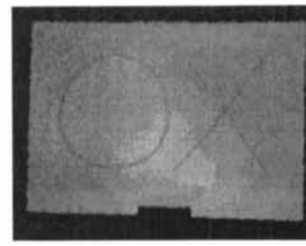
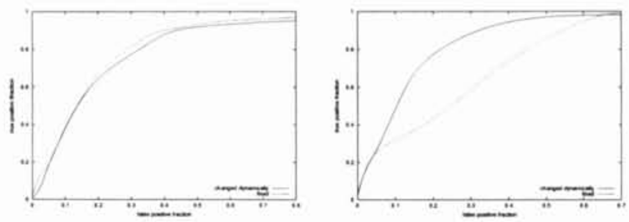


Figure 8: Plane object used as the input plane.



(a) Perpendicular plane (b) Slanted plane

Figure 9: ROC curves for two experiments.

3.4 Continuous Superimposed Display of the Input Locus

To correctly transform the fingertip locus between different observed images, we have to estimate the 3D geometric configuration between the camera and the input plane. This procedure, however, requires a complicated reconstruction of 3D information.

In the proposed system, therefore, the approximate transformation is employed: for the transformation between the current image and the image buffer for recording the fingertip locus, the following approximate equation is employed:

$$P(u,v) = (1-v)((1-u)P_{00}+uP_{10})+v((1-u)P_{01}+uP_{11}), \quad (1)$$

where $0 \leq u \leq 1$, $0 \leq v \leq 1$ and P_{00}, P_{01}, P_{10} and P_{11} denote a 2D point. This equation transforms a 2D point in the square (denoted by B_s) determined by $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$ to a 2D point in an arbitrary tetragon determined by P_{00}, P_{01}, P_{10} and P_{11}. We regard B_s and $(P_{00}, P_{01}, P_{10}, P_{11})$ as the image buffer and the coordinates of the four vertices of the input plane, respectively.

4 Experiments

We conducted experiments to verify that our proposed system works well. Our system consists of a PC (PentiumIII 1GHz), a wearable active-infrared camera and a HMD. The active-infrared camera system consists of a SONY XC-EI50 with an infrared-pass filter and 24 small infrared-LEDs. The size of a captured image is 640 × 480. With these resources, the system could capture images and process them at about 0.09[sec] intervals (11[frames/sec]) on average. In all the experiments, we used a sheet of A4-size paper shown in Fig.8 as the input plane. To verify the performance of the proposed system, reference figures (i.e., the circle and the cross) were drawn on the input plane object.

4.1 Correctness of the discrimination between input and non-input states

The user tracked the reference figure on the input plane with the fingertip. We assume that the ground truth of the drawn fingertip-locus was identical to the reference figure. If the distance between the position of the detected fingertip and the reference figure is smaller than 5[pixels] when the system regards the fingertip as touching the input plane, this detected position is considered to be the correct input data. Otherwise, this detected position is considered to be the error input data. The former and latter mean "the system correctly detects the input state when the user touches the input plane" and "the system incorrectly detects the input data when the user does not touch the input plane", respectively. To verify whether or not the system correctly discriminates between the input and non-input states, we evaluated the following two rates in the ROC curve:

True positive: The rate of correct inputs.

False positive: The rate of error inputs.

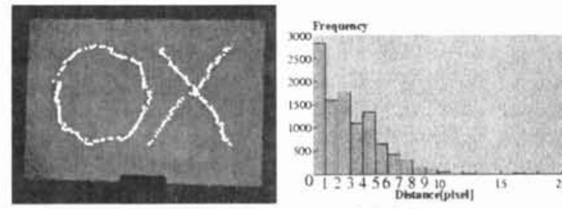
We evaluated these values for the proposed adaptive threshold method (the solid line in Fig.9) and the fixed threshold (the dotted line in Fig.9). To obtain the ROC curves, 1) in the fixed-threshold method, the threshold was adjusted, and 2) in the proposed method, line L in Fig.7 was translated along the vertical axis. Figures 9 (a) and (b) show the results for the plane perpendicular to the optical axis of the camera and the slanted plane, respectively. Especially for input in the slanted plane, the proposed method provided better results. We can, therefore, confirm that the proposed adaptive threshold method is required for a wearable computing environment.

4.2 Accuracy of the input data

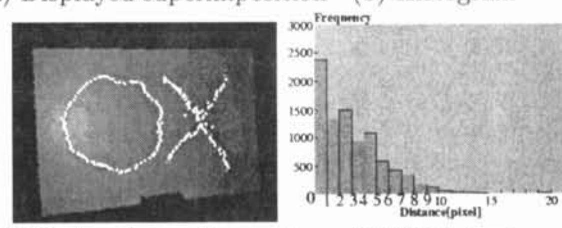
Next, we verified the distance accuracy of the input data by evaluating the error distance between the reference figure and the drawn locus. The user tracked the reference figure with the fingertip as well as in the above experiment. Figures 10 (a) and (c) show the drawn loci. Figures 10 (b) and (d) show the histograms whose horizontal and vertical axes indicate the error distance and the number of the drawn points. In the perpendicular plane, the average, median and variance values of the error distance were 3.06, 2.83 and 5.64, respectively. In the slanted plane, the results were 3.15, 2.83 and 6.15. In both cases, the rates of the drawn points, whose error distances were within 5[pixel], were over 90%. We consider this accuracy to be enough for practical use.

4.3 Effectiveness of the superimposed display

To verify the effectiveness of the superimposed display, we projected the reference figure from the perpendicular plane (shown in Fig.11 (a)) to the slanted plane (shown in Fig.11 (b)) through the buffer image. The projected figure did not completely coincide with the reference figure on the slanted plane. We could, however, continuously draw in the slanted input plane while visually checking the geometric configuration between the projected figure and the newly input locus.

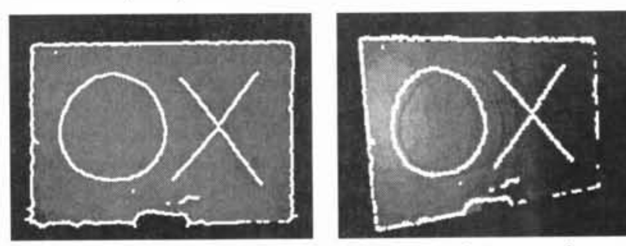


(a) Displayed superimposition (b) Histogram



(c) Displayed superimposition (b) Histogram

Figure 10: Input accuracy. Upper: Input plane perpendicular to the optical axis of the camera, Lower: Slanted input plane.



(a) Perpendicular plane (b) Slanted input plane

Figure 11: Superimposition while moving the plane.

5 Concluding Remarks

By employing the properties of the active-infrared camera, we developed the wearable virtual tablet. A user can draw an arbitrary locus on a rectangular plane held by the user's hand. Since the drawn locus is displayed on the HMD and its shape is dynamically adjusted depending on motions of the user and input plane, the user can continuously utilize the WVT in a wearable computing environment.

This research is supported by Core Research for Evolutional Science and Technology (CREST) Program "Advanced Media Technology for Everyday Living" of Japan Science and Technology Corporation.

References

[1] Y. Yamamoto and I. Shiio, "A Simple AR System for Casual Communication", in *Proc. of Workshop on Interactive Systems and Software (WISS 2000)*, pp.117–124, 2000. (Written in Japanese)

[2] Y. Muraoka and T. Sonoda, "A Letter Input System of HandWriting Gesture - A User Interface for Wearable Computers -", in *Proc of Interaction 2001*, pp.3–10, 2001. (Written in Japanese)

[3] H. Sasaki, T. Kuroda, Y. Manabe and K. Chihara, "Hand-Area Extraction by Sensor Fusion Using Two Cameras for Input Interface of Wearable Computers", in *Proc of The 12th Scandinavian Conference on Image Analysis*, pp.779–784, 2001.

[4] H.Koshimizu and M.Numada, "On a Basic Consideration of the Warp Model of Hough Transform", in *Proc of Machine Vision Application* (MVA '92), pp.7–9, 1992.