Bare Hand Interface for Interaction in the Video see-through HMD based Wearable AR Environment*

Taejin Ha and Woontack Woo

GIST U-VR Lab.
Gwangju 500-712, South Korea
{tha, wwoo}@gist.ac.kr

Abstract. In this paper, we propose a natural and intuitive bare hand interface for wearable augmented reality environment using the video see-through HMD. The proposed methodology automatically learned color distribution of the hand object through the template matching and tracking the hand objects by using the Meanshift algorithm under the dynamic background and moving camera. Furthermore, even though users are not wearing gloves, extracting of the hand object from arm is enabled by applying distance transform and using radius of palm. The fingertip points are extracted by convex hull processing and assigning constraint to the radius of palm area. Thus, users don't need attaching fiducial markers on fingertips. Moreover, we implemented several applications to demonstrate the usefulness of proposed algorithm. For example, "AR-Memo" can help user to memo in the real environment by using a virtual pen which is augmented on the user's finger, and user can also see the saved memo on his/her palm by augmenting it while moving around anywhere. Finally, we experimented performance and did usability studies.

1 Introduction

The vision based hand interface enables more natural and intuitive interaction but it has inherent problems, as follows: First, it is difficult to construct the hand color model when the hand object position is unknown. Thus, it requires the user to be provided additional input events (e.g. mouse button) to learn the hand color model in specific area. In the conventional systems, users wore color gloves to easily extract the hand object from arm where the hand object has similar color distribution as the arm [8]. In this case, user may feel inconvenience. Lastly, the users usually attach the fiducial markers [1] on the fingertips to detect fingertip points as well but it is also unnatural and inconvenient to use.

Therefore, we proposed the following approaches to overcome aforementioned problems. The proposed system composed of three sequential stages, first, detect the hand object from input image (e.g. region of interest) by exploiting the template matching based on invariant features (e.g. Hu moments [2]) and then learn the

* This research is supported by the UCN Project, the MIC 21C Frontier R&D Program in Korea and GIST ICRC & CTRC.

detected object's color distribution automatically. It is possible to the extract hand object from arm, even though users are not wearing gloves so they can feel easy. Then, extract the hand object by using the distance transform [3] and radius of palm. Finally, extract the fingertip points by finding convex hull points of the hand object and constraining radius of palm.

We implemented several applications based on the proposed technique. The system "AR-Memo" can help user to memo in the real environment by using a virtual pen, and the user can also see the saved memo on the palm while moving around. In addition, the user can change the color or the thickness of a pen through "AR-Menu" and the user can play "AR-Tetris" game by using two hands. Finally, the proposed system has been evaluated regarding the performance and usability test.

This paper is composed as follows. In section 2, we explain the algorithm and implementation. In section 3, we address application. In section 4, we show the experimentation results and conclusion in last section.

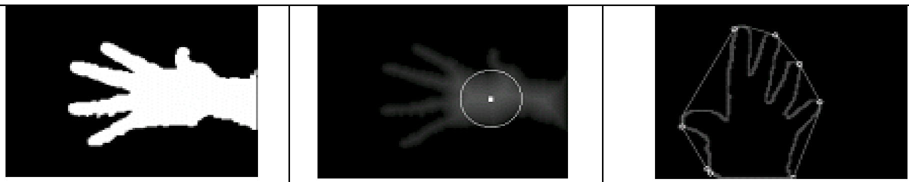
2 Algorithms for the vision based hand interface and application

Various methodologies have been proposed to learn color distribution of the hand object by using the user's hand on the certain area of a screen and pushing mouse buttons [9]. Most of the conventional methodologies require additional user input so it is uncomfortable. The proposed method automatically learns color distribution of the hand object in consideration of HSV color space. A binary image is obtained based on the intensity of center point of the interest region. After contour processing, the maximum area of object and Hu moment are calculated. If the distance between the detected object and the known hand template object is below threshold, then the hand color distribution is constructed through the three sigma rule.

Next, we try to extract the hand object from the arm, because the hand and arm have the same color. It is very difficult that the size of the hand and its direction change continuously. In this paper, we applied the distance transform to silhouette image of the hand object. The highest intensity value in the silhouette image indicated the center point of a palm area. The size of radius of palm can be obtained by inspecting the intensity value $I_{(x, y)}$ of point is equal to the zero. In order to extract fingertip points, we used "3 coins algorithm" [5]. However, there are too many convex hull points are existed so we consider the constraint that the distance between convex hull points for each point on the fingertips. We set constraint that the points above 1.5 times of palm's radius.

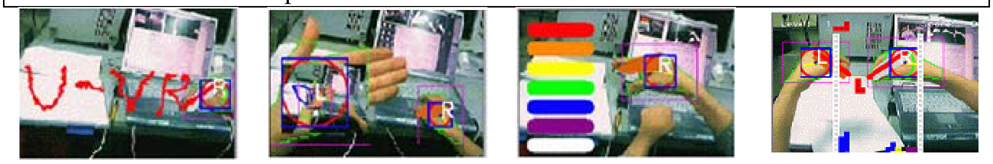
After finding the center point and the radius of a palm, the search window is set to reduce the redundant image processing. The mean shift algorithm [4] can be used to track the hand object but the size of search window expands continuously that contain the similar color object (e.g. face or hands). In the proposed approach, the center point of the search window is set as the center point of palm area and the size of search window is set based on the radius of palm.

We implemented several applications based on the proposed technique.



(a) (b) (d)

Fig. 1. (a) Silhouette image of a hand object (b) the center point and radius of palm (b) extracted convex hull points



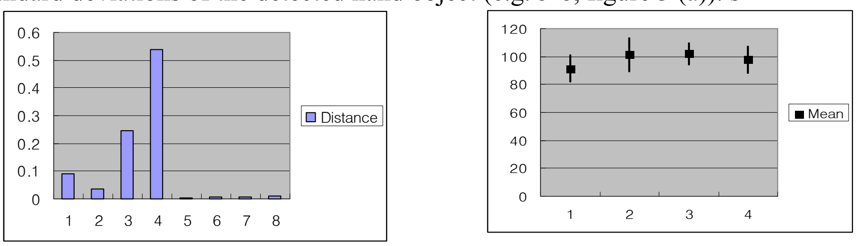
(a) (b) (c) (d)

Fig. 2. (a) Memo by using a virtual pen (b) a saved image is augmented on the palm. (c) Change color of a pen (d) the user can move blocks using the finger tips in the Tetris game

4 Performance comparison and usability test

The proposed scheme has been coded with OpenCV library [7] and implemented under mobile computer equipped with Intel mobile processor 1.13 GHz and 512 RAM. We used a cheap USB camera; resolution is 320*240(pixels), frame rate 25f/s and HMD, i-visor DH-4400VP [6]. The experiment is executed in the usual indoor environment.

Through the experimental experience, we decided the threshold for template matching based on the distance value. If the distance value is less than 0.01, then we could properly detect the hand object. Fig. 3 (b) represents the mean values and its standard deviations of the detected hand object (e.g. 5-8, figure 3 (a)). s



(a) (b)

Fig. 3. (a) The distance between segmented object and the template hand object. 1-4: false detection, 5-8: true detection, (b) Mean and standard deviation value of the hand object

Next experiment is about detection rate and usability test in two cases. We measured how many times hand object is detected in 5 seconds. When the users wear the gloves, the number of detection per second is 13.56 in an average. In the bare hands case, 11.7 counts per second and 87.5 % of subjects prefer usage of the bare hand more than wearing gloves (6.25 %) and etc are 6.25 %. Also, when fiducial markers are attached on fingertips, the number of detection per second is 14.26 whereas the bare hand case, 10.3 per second. 81.25% of subjects prefer the usage of bare finger more than using fiducial markers (6.25 %) and etc is 12.5 %.

5 Conclusion and Future works

In this paper, we proposed a vision based bare hand interface for interaction in augmented reality environment. The proposed methodology automatically learned color distribution of the hand object and tracking under the moving camera. Also, we extracted the hand object from arm but not using the glove. We also detected the fingertip points but not using the fiducial markers. We implement several applications.

Through experiments, we found the users prefer interaction using the bare hand and the bare finger rather than wearing the glove or attaching marker on user's fingertips, even though performance is little bit dropped. We believed that the vision based bare hand interface will be a natural and intuitive user interface.

As future works, we are now considering to applying estimator for more robust augmentation and designing the hand segmentation model under various lighting condition.

6 References

[1] ARToolKit, http://www.hitl.washington.edu/artoolkit/

[2] Hu, M.K., "Visual pattern recognition by moment invariants", IEEE Transactions on Information Theory, IT-8, pp 179-187, 1962

[3] Shapiro, L., and Stockman, G., Computer Vision, 196-197, Prentice Hall, 2001

[4] Gary R. Bradski, Computer Vision Face Tracking For Use in a Perceptual User Interface, Microcomputer Research Lab, Santa Clara, CA, Intel Corporation

[5] Avis, D. and Toussaint. "An optimal algorithm for determining the visibility of a polygon from an edge", IEEE Trans. Comp. C-30 (1981), 910-914

[6] Daeyang i-visor DH-4400VP, http://www.mpcclub.ru/index.php?action=product&id=3340

[7] Open Source Computer
Vision Library, www.intel.com/research/mrl/research/opencv

[8] Ross Smith, Wayne Piekarski, and Grant Wigley, "Hand Tracking For Low Powered Mobile AR User Interfaces", In 6th Australasian User Interface Conference, Newcastle, NSW, Jan 2005

[9] Cristina Manresa, et al. "Hand Tracking and Gesture Recognition for Human-Computer Interaction", Electronic Letters on Computer Vision and Image Analysis 5(3): 96-104, 2005