A Novel Projector-Camera Interaction System with the Fingertip

Qun Wang[1,2], Jun Cheng[1,2,3], and Jianxin Pang[1,2]
[1]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[2]The Chinese University of Hong Kong, Hong Kong, China
[3]Guangdong Provincial Key Laboratory of Robotics and Intelligent System
Email: qun.wang@siat.ac.cn, jun.cheng@siat.ac.cn, jx.pang@siat.ac.cn

***Abstract*—In this paper, we propose a vision based human computer interaction (HCI) system, and user can interact with the computer by using the fingertip. And a novel algorithm is proposed, which robustly detects fingertip and archives touch-operated controls with a high precision of few millimeters. The fingertip detection algorithm is developed based on the prediction of the contents which is projected by projector, and a binocular vision system based on two cameras is applied for detecting the depth of fingertips and touch operation. The experiment shows good results and good performance. We have two important contributions: first, we develop a simple and robust hand detection method which can work well in a wide variety of situations including dark and luminous ambient illumination. Second, the two cameras comprise a stereo vision system. We determine whether a physical touch takes places by triangulation.**

***Index Terms*—hands detection, fingertip detection, stereo vision, touch detection, touch-operated controls**

I. Introduction

During the past few years, the technology of camera and projector are getting great progress. The touchscreen provides a fantastic user experience that the user can interact with computer and mobile terminal equipments using simple gestures by touching the screen with one or more fingers. Compared with the input-output user interface of traditional computers, it is more natural that using our fingers to drag items on the virtual screen of the computer, to open files and folders, to scroll pages and so on. In this paper our motivation aims on converting any plain surface into a touchscreen.

The proposed system includes a projector and a binocular vision system based on two cameras, and the projector projects the contents on the wall or any other plain surface, then the cameras capture the projected ones. We realize human computer interaction with the help of computer vision. There are two challenges in the system; the first challenge is how to extract human hand from the camera image and how to detect the finger tip, and the complexity of the projected contents make that more difficult. The second challenge is how to recognize whether a physical touch takes places，which achieves the touch-operated control..

Vision and finger-based human computer interaction is an ongoing research area, and there has been significant research carried out to make human computer interaction friendly and intuitive. Letessier [1] presented a system that tracks the 2D position of the tips of bare fingers on a planar display surface, but he neglected finger clicking detection. In [2], the authors proposed a hand gesture-based human computer interaction system, where users can interact with the projected screen using his fingertips which are tracked in air by the camera using ‘Camshift’ tracker. They use fingertip path for finger clicking detection. In [3], they present real-time techniques for recognizing “touch” and “point” gestures on steerable projected displays produced by a new device, Button touches are detected by examining the hand trajectory for several specific patterns. All of the above, click event is determined through a delay based scheme, which has limited usability in applications which require fast response. Ankur [4] described a machine learning algorithm that can identify fingertips and detect touch with a precision of a few millimeters above the surface using a pair of cameras mounted above the surface. All of the above methods, hand detection methods are developed on both the color and shape of the hand. Hand color is changing with the projected image, even sometimes the projected image has a human hand. Jingwen Dai [5] presented a hand detection method using the information that the computer knows the projected contents. They use imperceptible structured light for touch detection, with the entire system comprising a projector and a camera. But the imperceptible structured light require high rate camera, which has limited the applications.

Our proposed system aims at changing any plain surface into a touchscreen. We have two important contributions: first, we develop a simple and robust hand detection method which can work well in a wide variety of situations including dark and luminous ambient illumination. It can also work well in a dynamic background. Second, the two cameras comprise a stereo vision system. We determine whether a physical touch takes places by triangulation.

Manuscript received February 11, 2013; revised April 16, 2013.

doi: 10.12720/joig.1.2.80-84

The block diagram of the system is shown in Fig. 1. The remainder of this paper is structured as follows. In the next section, we propose our robust hand detection method. In Section III, the strategy of fingertip detection method is described. In Section IV, we introduce our method of compute the distant between the display surface and fingertip, and then determine whether a physical touch takes places. In Section V, system setup and experimental results are shown. Conclusion and future work are offered in Section VI.

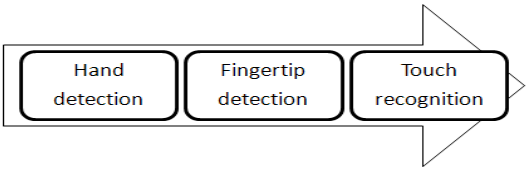

Figure 1. Block diagram of the system.

II. HAND DETECTION

There are many methods in hand detection. Most of them are based on color and shape of the hand. In this paper, the pro-cam system allows us to know where the video contents are projected and how they should appear in the image data. We can use this information to predict an image that the camera should read, and analysis the difference between the predicted image and the read image to detect human hands. In order to achieve this goal, we need an accurate prediction of the appearance of the computer projected content as viewed by the camera. This requires two basic components:

- *Geometric calibration:* It concerns the mapping between the position in the camera view and the position in the projector screen.
- *Photometric Calibration:* It concerns the mapping between the actual color of the projected content and that seen by the camera.

A. Geometric Calibration

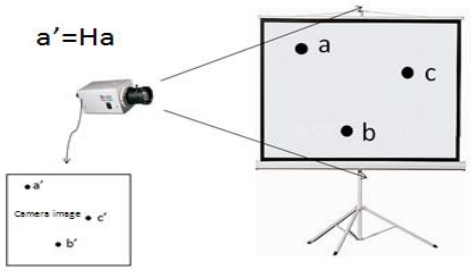

Figure 2. Geometric calibration

In order to predict camera image, we need to know the relationship between the pixels in the camera view and the pixels in the projector screen. This is the task of geometric calibration. The mapping between a point in the camera view and a point in the projector screen can be described by a 3×3 matrix H . Fig. 2 shows the relation between the two image planes. The projector can project the patterns we want, so we can project a chessboard and capture the image with camera. Then detect the corners of the chessboard. After this we can estimate the holography between the projector screen and the image plan of the camera using the detected corners and their corresponding known positions in the projector. As shown in Fig. 3.

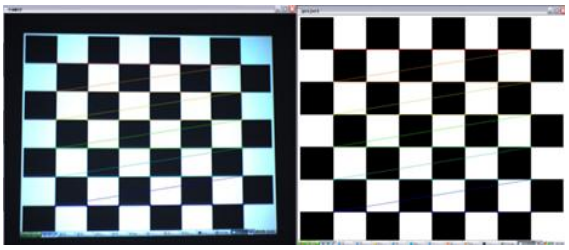

Figure 3. The chessboard in camera image and in projector screen

B. Photometric Calibration

In order to predict the image captured by the camera, for a given pixel in the projector space, we know its corresponding position in the camera image through geometric calibration; furthermore, we need to know what the corresponding color should look like in the captured image, and this is the task of photometric calibration. Due to different color spectra responses of the two devices, the same color in the project space appears very differently in the camera space. The issue of color non-uniformity in the projected image further complicates this matter. The same input color, when displayed in the projector's corner, is different from that in the projector's center. Therefore, photometric calibration should be both color- and position-dependent [6]. In X. Chen's method [7], the photometric model of a projector-camera system at each pixel can be written as:

$$C = A(VP + F) \tag{1}$$

where vector C is the camera captured brightness. The Vector P is projector brightness. The matrix A is reflectance of the surface. The vector F shows the contribution of ambient light. The matrix V is called color mixing matrix which describes the interaction of color channels of a projector-camera system. Our goal is to compute the A, V and F .The following steps describe how to get the A, V and F . For each pixel in the projector using H establish correspondence between the camera view and the projector screen.

1). *Project a black image to compute* F
2). *Project pure red, pure green and pure blue image to compute the color mixing matrix* V .
3). *Project a white image to compute vector* A.

Given arbitrary display content, we could generate the predict image by (1). Fig. 4 (a) and (b) shows the projected image and the predict image.

C. Hand Detection

Now we get the predict image, the next step is hand detection. When we interact with the computer on the display surface, we change the surface albedo of the display surface. Therefore extracting human hand relates to detecting the changes on the surface albedo.

Assuming Q is the incident light, A is the surface albedo of the display surface, T is the pixel-wise color

transformation due the camera sensor, and C is the camera captured brightness, we have $C = A \times T \times Q$.[8] If nothing before the display surface, the captured image I should be equal to C . If there is a hand before the screen, the surface albedo changes, as denoted by A' . The captured image can then be described $I = A' \times T \times Q$. We can compute the albedo change by estimating the albedo ratio $a = A'/A$ of the pixel [x, y] in color channel $c \in \{R, G, B\}$, which is given by

$$a_{[x,y,c]} = \frac{A'}{A} = \frac{I_{[x,y,c]}}{C_{[x,y,c]}} \tag{2}$$

Based on the albedo ratio a , we can detect the hand. The albedo of the display surface region without hand should be close to 1. For a pixel[x, y], assuming the sum of the three channel albedo ratio is $a_{[x,y,sum]}$. For an image the average of the total three channel albedo ratio is $a_{[sum]}$. We use following decision rule:

Pixel [x, y] belongs to the hand region if and only if

$$a_{[x,y,R]} + a_{[x,y,G]} + a_{[x,y,B]} < s \times a_{[sum]}$$

or

$$s \times (a_{[x,y,R]} + a_{[x,y,G]} + a_{[x,y,B]}) > a_{[sum]} \tag{3}$$

where s is a tolerant scale of the albedo change, with typical value of 0.5~0.8. Fig. 4 shows the projected image, predict image, camera read image, and hand detection result.

III. Fingetip Detection

After detecting the hand region we employ curvature [2] for the segmentation of fingertips. The curvature of a point Pi of a contour is computed as

$$K(P_i) = \frac{P_iP_{i-x} * P_iP_{i+x}}{\| P_iP_{i-x} \| * \| P_iP_{i+x} \|} \tag{4}$$

where P_i is the point under curvature test, P_{i-x} is the preceding and P_{i+x} is the succeeding point on the contour and x is the displacement index. The typical value of x is 5. Below are the main steps:

1) Find the contour of the hand
2) For every point in the contour, compute its curvature K , if K > 0.1 then the point is considered as a candidate for fingertip.
3) Compute the center of mass of the contour.
4) Compute the distance of the candidate points to the center of mass, and then find the most distant point as fingertip.

Fig. 5 shows the fingertip detection results.

IV. Touch Recogenition

After fingertip detection, the next work is to examine if the fingertip touches the display surface. Two cameras comprise a stereo vision system, we can compute the distance between the fingertip and the display surfaces, and then recognize the touch action.

Accuracy and simplicity of system calibration is a key challenge in 3D computer vision tasks. Zhang's method [9], using a simple planar pattern has provided the research community with both easy-to-use and accurate algorithm for obtaining both intrinsic and extrinsic camera parameters. This algorithm was implemented in Matlab Camera Calibration Toolbox [10].

After system calibration, we can compute the distance D_1 between the fingertip to the display surface and the distance D_2 between the display surface to the camera with the help of triangulation. We use the follows decision rule:

A physical touch takes place if and only if

$$\frac{D_1}{D_2} < L \tag{5}$$

where L is a relative scale of the two distances, the typical value of L is 0.05.

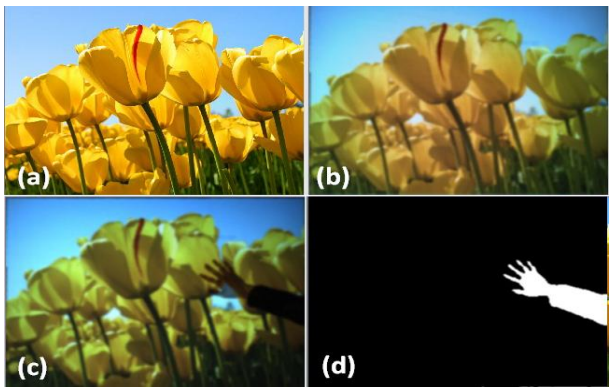

Figure 4. (a) Projected image, (b) Predicted image, (c) Read image, (d) Hand region.

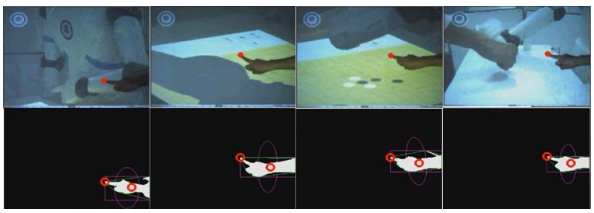

Figure 5. Fingertip detection results

V. Experimental Results

In order to test the system, we conducted experiments to evaluate fingertip detection and touch detection accuracy respectively.

For fingertip detection, we used various projectors (including large projectors with resolution of 1440×1050 and DLP Pico Projector with resolution of 640×480) and various video cameras (including industrial cameras and web cameras), under both artificial lighting and natural lighting conditions. The experiment shows good results and good performance. Fig. 6 shows the experiments results, the fingertip is marked with a red point. 6(a) is in natural lighting condition with white background, 6(b) is in natural lighting condition with black background, 6(c)

is in artificial lighting condition, 6(d) shows that in dark situation and black backgurond, we can hardly see the human hand, but our method still working well. 6(e) and 6(f) shows that although the projected content included a human hand(left), it doesn't impact our system.

For touch recognition, we used a SONY projector with resolution of 1440×1050 and two web cameras (30fps, resolution of 640×480). The distance between projection screen and the projector was 85cm.

The cameras were close to the projector. We used Microsoft Paint to test the performance of our system. The average error is within 0.7cm. Fig. 7 shows that we could paint on the projection screen with a fingertip.

For human computer interaction system, real-time performance is of great importance. We used a SONY projector with resolution of 1440×1050 and two web cameras (30fps,resolution of 640×480) and a computer with Intel Core2 Duo 3.20GHz CPU. The total time consumption is less than 20ms, indicating the system meets the requirement of real-time application.

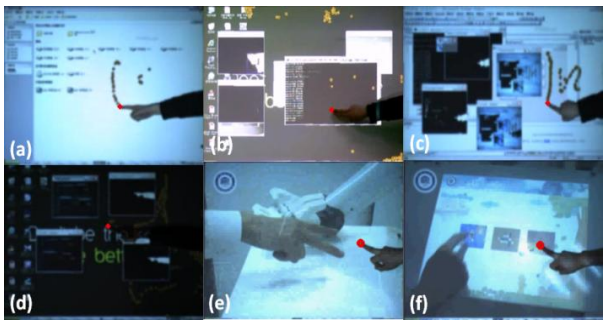

Figure 6. Fingertip detection results. The fingertip is marked with a red point. (d)In the dark situation, we can hardly see the hand, but our method still working well. In (e) and (f) the projected image includes a human hand (left) , but this doesn't impact our system.

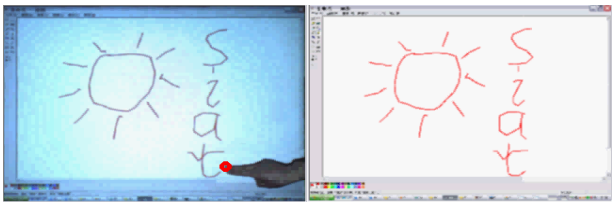

Figure 7. Touch recognition result, left figure shows camera captured image, right figure shows projected image.

VI. Conclusions And Future Work

We proposed a fingertip based human computer interaction system that provides a very intuitive and natural way to interact with computer.

The hand is first segmented using the "prediction method", then fingertips are located on the contour of the segmented hand and touch is recognized by the triangulation. Encouraging results are produced under various conditions and various backgrounds. The accuracy of the touch recognition is about 0.6 cm. We are looking to improve the touch recognition accuracy. The performance of the algorithm falls in very dark lighting conditions with black background and is also susceptible to reflections on the screen surface. Although vision based systems are often associated with such drawbacks, resolving these issues will be the focus of our future work.

Acknowledgment

Special thanks to CAS and Locality Cooperation Projects (ZNGZ-2011-012), Guangdong-HongKong Technology Cooperation Funding (2011A09- 1200001), Guangdong Innovative Research Team Program (No.201001D0104648280), Shenzhen Technology Project (ZD200904300074A).

References

[1] J. Letessier and F. Bérard, "Visual tracking of bare fingers for interactive surfaces," in *Proc. 17th Annual ACM Symposium on User Interface Software and Technology*, 2004, pp. 119-122.

[2] S. A. H. Shah, A. Ahmed, I. Mahmood, and K. Khurshid, "Hand gesture based user interface for computer using a camera and projector,"in *Proc. IEEE International Conference Signal and Image Processing Applications*, 2011, pp. 168-173.

[3] R. Kjeldsen, C. Pinhanez, G. Pingali, J. Hartman, *et al.*, "Interacting with steerable projected displays," in *Proc. Fifth IEEE International Conference Automatic Face and Gesture Recognition*, 2002, pp. 402-407.

[4] A. Agarwal, S. Izadi, M. Chandraker, and A. Blake, "High precision multi-touch sensing on surfaces using overhead cameras," in *Proc. Second Annual IEEE International Workshop Horizontal Interactive Human-Computer Systems*, 2007, pp. 197-200.

[5] J. Dai and R. Chung, "Making any planar surface into a touch-sensitive display by a mere projector and camera," in *Proc. IEEE Computer Society Conference Computer Vision and Pattern Recognition Workshops*, 2012, pp. 35-42.

[6] M. Liao, R. Yang, and Z. Zhang, "Robust and accurate visual echo cancelation in a full-duplex projector-camera system," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 1831-1840, 2008.

[7] X. Chen, X. Yang, S. Xiao, and M. Li, "Color mixing property of a projector-camera system," in *Proc. 5th ACM/IEEE International Workshop on Projector Camera Systems*, 2008, pp. 14.

[8] M. Zhou, Z. Zhang, and T. Huang, "Visual echo cancellation in a projector-camera-whiteboard system," in *Proc. International Conference Image Processing*, 2004, pp. 2885-2888.

[9] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1330-1334, 2000.

[10] J. Y. Bouguet. Camera Calibration Toolbox for Matlab. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc

Qun Wang received the bachelor degree in microelectronics form the Nankai University in Tianjin (2010). He is currently a graduate student of Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, majoring in computer science. His research interests include pattern recognition, computer vision, and machine learning.

Jun Cheng received the B.Eng., B.Fin., and M.Eng. Degrees from the University of Science and Technology of China, Beijing, in 1999 and 2002 respectively, and the Ph.D. degree from the Chinese University of Hong Kong in 2006. Currently he is with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, as a Professor and Director of the Laboratory for Human Machine Control. His research interests include Computer Vision, Pattern Recognition, Human Machine Interface and Robotics.

Jianxin Pang is a senior engineer at Shenzhen Institutes of Advanced Technology of Chinese Academy of Science. He received the B.Eng. degree (2002) and the Ph.D. degree (2008) from the University of Science and Technology of China (USTC). His research interests include image and video understanding, Human Computer Interaction, visual quality assessment, and computational vision.

