Acquire User's Distance by Face Detection

Cheng-Yuan Ko, and Liang-Gee Chen*, *Fellow, IEEE,*
Graduate Institute of Electronics Engineering, National Taiwan University, Taipei, Taiwan, R.O.C

***Abstract*--In this paper, we propose an algorithm using only two commodity webcams without calibration to detect distance between user and display by face detection. According to the experimental result, the correlation coefficient of distance and the reciprocal of disparity is up to 0.9984.**

Our proposed algorithm can provide user's depth information with high accuracy from calibration free stereo capture image pair for the application, such as interactive 3DTV or user-aware auto-stereoscopic display.

Index Terms—calibration-free, depth, distance estimation

I. INTRODUCTION

In recent years, how to interact with device or machine is a very active research area in Human-Computer Interaction (HCI). To achieve this goal, the most important information should be considered is user's location. There are two kinds of systems are popular with researchers. One is sensor based, such as IR-based camera [1]. Although IR-based camera can provides a convenient, fairly accurate depth measurement. However, due to the IR-based camera detects the user real time action with the infrared through a built-in camera sending active laser, the interference issue is a big problem.

Another method is using stereo camera. Using stereo camera to capture the left view and right view simultaneously, and then do the stereo matching process to get depth map [2]. According the depth map, we can get the user's depth information easily.

However, before the stereo matching process, we must have to calibrate stereo camera. So the previous work proposed Mask-based stereo matching algorithm to get the user's location and calibration-free [3]. The limitation of Mask-based stereo matching algorithm is that when user is too close to stereo camera, the illumination variance in the capture images will be very obvious. So the left mask and right mask which are generated by background subtraction will have very poor quality.

Lee Chong Wan et al. proposed a skin color based stereo vision tracking system [4]. They also do the depth estimation by skin segmentation result, but their working range only from 15 cm to 65 cm and the accuracy approximately 80%.

In this paper, we proposed a face detection-based stereo matching algorithm to overcome the situation which has large illumination variance. And the proposed algorithm can use uncalibrated input captures. .

The rest of this paper is organized as follows: Section II presents our proposed face detection-based algorithm. Section III presents the experiment setup and results. Section IV presents our conclusions.

II. PROPOSED ALGORITHM

We will present the details of the face detection-based stereo matching algorithm in the following. The overall system flow is shown as Fig 1.

A. Face detection

Here, we solved a problem about face detection in this paper is the false detection problem. We use Haar-like feature classifier in OPENCV to detect user's face. Though, we can set lower bound value of the face width and height. However, the size of user's face will be different due to the location of the user. When user's location is nearer to the camera, the user's face is bigger; when user's location is far away the camera, the user's face is smaller. So we can only set the lower bound with a small value to guarantee the system can detect the user's face in every distance. In this condition, false detection may be occurred, as shown as Fig. 3. Fig. 3(b)(e) shows case (I) which occur the false detection in both left view and right view. For this case, we find each face detection results pairs in left view and right view, and then do the face detection-based stereo matching. According to the geometric relationship with the size of face and the distance of user, if the size difference between the face detection results and the size which is estimated by geometric relationship is too large, then we considered it as a false detection. Finally, we can rule out those false detections easily, as shown as Fig. 3(c)(f).

Another case is that the false detection only occurred at one view, case (II) is shown as Fig. 4(b)(e). First, we do the same flow of case I, but when some detection results cannot find out the correspondent in another view, we ignore it, then the results are shown as Fig. 4(c)(f).

B. Face detection-based stereo matching

In this part, we only use formula 1 to calculate the disparity for user's face.

$$\text{Disparity} = |Lx - Rx| \quad (1)$$

Where Lx is the center of user's face in left view and Rx is the center of user's face in right view. After we get the disparity, we transform the disparity to the real distance between the user and the camera.

III. EXPERIMENTAL RESULTS

In this section, we present our experimental result and measure the linearity of actual physical distance versus the reciprocal of disparity as the basis to judge the accuracy. In our experiment, we use two commodity webcams to build a

stereo vision system. The resolution of input images for left view and right view are both VGA(640x480). In the beginning, the user sits in front of cameras which the distance between user and cameras is 30 cm. Then we increase the distance between user and cameras incrementally up to 150 cm every 30 cm. The results are show as Fig. 2 and TABLE II.

Fig. 2 showed the observations, the correlation coefficient of distance and the reciprocal of disparity is up to 0.9984. It is worth mentioning that the advantage of the proposed method is that we can get very good results as long as the difference of horizontal level of two cameras are not too much.

We also make a comparison with Lee Chong Wan et al.[3] as shown in Table I. The formula for calculate accuracy is shown as follow:

$$\text{Accuracy} = \frac{|\ \text{Real distance} - \text{Estimated distance}\ |}{\text{Real distance}} \quad (2)$$

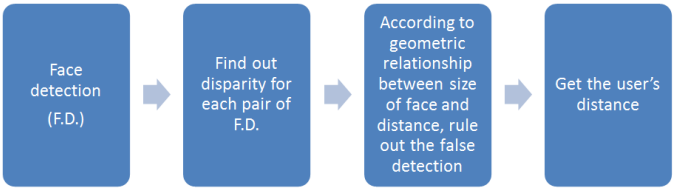

Fig. 1. The overall system flow for the proposed algorithm

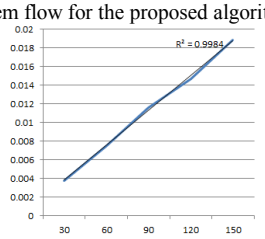

Fig. 2. Linearity of actual physical distance and the reciprocal of disparity. Y-axis represents the reciprocal of disparity, X-axis represents distance between user and cameras.

IV. CONCLUSION

We proposed a face detection-based stereo matching algorithm to detect the user's distance only by face detection. Our experimental result shows that the correlation coefficient of distance and the reciprocal of disparity is up to 0.9984.

Because of traditional stereo system using stereo matching to find the user's depth by calibrated input image, the advantage of this algorithm is that we can get very good results with low complexity and calibration free.

Our proposed algorithm can provide user's depth information with high accuracy from calibration free stereo capture image pair for the application, such as interactive 3DTV or user-aware auto-stereoscopic display.

TABLE I
COMPARISON

Algorithm	Operational range(cm)	Accuracy
Lee Chong Wan et al.[3]	15~65	80%
Proposed	30~150	98.7%

TABLE II
REAL DISTANCE V.S. ESTIMATED DISTANCE

Real distance(cm)	Estimated distance(cm)	Standard deviation(cm)
50	50.76086416	2.126539805
60	61.24376584	1.455897765
70	67.58129317	2.551406558
80	78.14902564	1.850974359
90	88.48688889	1.934786336
100	101.408682	1.981425719

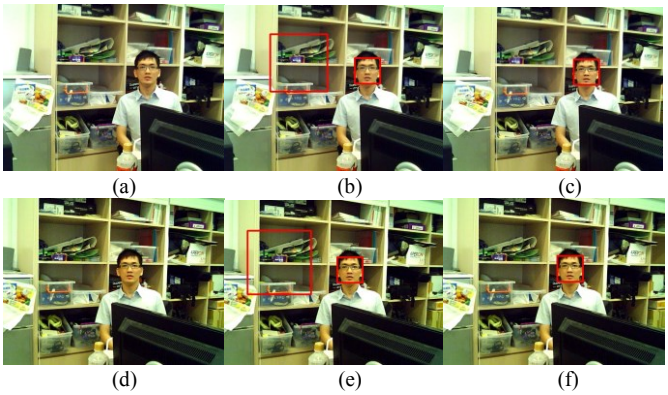

Fig. 3. Case (I) for false detection in both left view and right view. (a) left view image (b) face detection result with false detection in left view (c) face detection result without false detection in left view (d) right view image (e) face detection result with false detection in right view (f) face detection result without false detection in right view

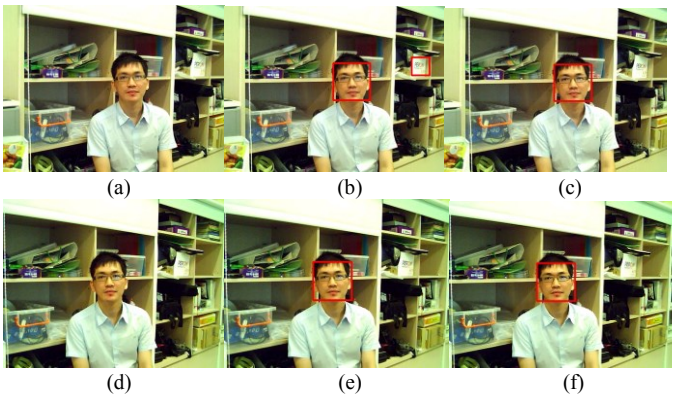

Fig. 4. Case (II) for false detection in only one view. (a) left view image (b) face detection result with false detection in left view (c) face detection result without false detection in left view (d) right view image (e),(f) face detection result without false detection in right view

REFERENCES

[1] Oreifej, Omar, and Zicheng Liu, "HON4D: Histogram of Oriented 4D Normals for Activity Recognition from Depth Sequences," *Computer Vision and Pattern Recognition (CVPR),* June 2013.

[2] Cheng-Yuan Ko, Chung-Te Li, Chen-Han Chung, and Liang-Gee Chen, "3D hand localization by low-cost webcams," *IS&T/SPIE Electronic Imaging (IS&T/SPIE EI),* Jan. 2013

[3] Cheng-Yuan Ko, Chung-Te Li, Chien Wu, and Liang-Gee Chen, "An Efficient Method for Extracting the Depth Data from the User," in *International Conference on 3D systems and Applications (3DSA),* June 2012.

[4] Lee Chong Wan, Patrick Sebastian, and Yap Voon, "Stereo Vision Tracking System," in *International Conference on Future Computer and Communication (ICFCC), Kuala Lumpur, Malaysia*, Apr. 2009