Proceedings of the 2005 IEEE
Engineering in Medicine and Biology 27th Annual Conference
Shanghai, China, September 1-4, 2005

Webcam Mouse Using Face and Eye Tracking in Various Illumination Environments

[1] Yuan-Pin Lin, [1] Yi-Ping Chao, [2] Chung-Chih Lin, and [1] Jyh-Horng Chen
[1] *Institute of Electrical Engineering, National Taiwan University*
[2] *Department of Computer Science and Information Engineering, Chang Gung University*

***Abstract*— Nowadays, due to enhancement of computer performance and popular usage of webcam devices, it has become possible to acquire users' gestures for the human-computer-interface with PC via webcam. However, the effects of illumination variation would dramatically decrease the stability and accuracy of skin-based face tracking system; especially for a notebook or portable platform. In this study we present an effective illumination recognition technique, combining K-Nearest Neighbor classifier and adaptive skin model, to realize the real-time tracking system. We have demonstrated that the accuracy of face detection based on the KNN classifier is higher than 92% in various illumination environments. In real-time implementation, the system successfully tracks user face and eyes features at 15 fps under standard notebook platforms. Although KNN classifier only initiates five environments at preliminary stage, the system permits users to define and add their favorite environments to KNN for computer access. Eventually, based on this efficient tracking algorithm, we have developed a "Webcam Mouse" system to control the PC cursor using face and eye tracking. Preliminary studies in "point and click" style PC web games also shows promising applications in consumer electronic markets in the future.**

I. INTRODUCTION

THE disabled who have limited voluntary motion often suffer great difficulties to access computer for internet, media entertainments, and information searching, etc. Hence, assistive technology devices have been developed to help the disabled. However, most of them require extra instruments which are uncomfortable to wear or use, such as infrared appliance, electrodes, goggles, and mouth sticks, etc.

Recently, due to the enhancement of computer performance and popular use of webcam, it is common to construct Human-Computer-Interaction (HCI) systems by using novel and intelligent video processing and computer vision techniques. These systems track the users' gestures with a video camera and translate them into the displacement of the mouse cursor on the screen. The "Camera Mouse" system has been developed to provide computer access for people with severe disabilities, which allows them to interact with their environment by spelling out message and exploring [3]. But drawback of the system needs another vision computer to execute the visual tracking task. Besides, the system involved expensive hardware such as video captured board, data acquisition board, and PTZ camera. In [2], the authors built an intelligent hand-free input device which allows users use their nose as a mouse based on visual tracking of face and nose with two low resolution cameras.

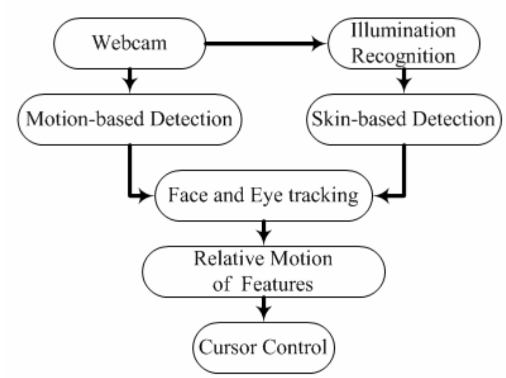

Fig. 1. Face tracking block diagram.

According to [4], a survey in visual tracking of face, human skin color has been used and proven to be an effective feature in many applications. To overcome the variation of lighting conditions would be the major task in skin-based face detection. The skin-tone range varies with different environments, so that stability and precision of the system would decrease dramatically[1]. For this reason, nonlinear color transform [5] and color histogram normalization [6] have been proposed to eliminate the lighting effects. But the computational complexity would increase the CPU loading in real-time implementation. In this study we establish an illumination-independent system combining illumination recognition method and adaptive skin models to obtain face tracking task. Besides, we use irises' characteristic, exhibiting low intensity of luminance component, to identify the eye features. Then, based on relative motion of face and eye features, the mouse cursor on the screen can be controlled by head rotation of users. To compare previous works, the goal of our research is to develop friendly and robust "Webcam Mouse" system via single webcam, and the system block diagram is shown in Fig. 1.

The organization of this paper is the following. The methodologies of face tracking, eye tracking, and mouse control strategy are detail described in Section 2. Section 3 presents the results of system implementation. Finally, we show some discussions and conclusions in Section 4 and Section 5 respectively.

II. METHODOLOGY

A. Skin-tone Color Distribution

Traditionally, the format of digital images and video sequences is RGB color model which separates effects of illumination fluctuation into R, G, and B color bands uniformly. For maintaining the stability of skin-based face tracker, this drawback leads to complicated algorithm to eliminate the fluctuation effects. In contrast with RGB model, YC_bC_r model distinguishes luminance component (Y) and chrominance component (C_b and C_r) independently. This advantage would be more suitable to decrease the luminance variation. Therefore, due to the consideration of illumination effects and computation cost, we only adopt the C_b and C_r components to construct the face tracking algorithm in this study.

In order to evaluate the feasibility of face tracking in C_r-C_b subspace, we analyzed a set of facial pixels on images captured via webcam with different subjects, head poses, and environments. Four results from analysis are in the following: (1) the distributions of face ROI and background are clustered appearance in C_r-C_b subspace; (2) skin-tone distribution is invariant to head poses in C_r-C_b subspace; (3) skin-tone distributions of different subjects but with the same racial are similar under the same environment; (4) skin-tone distribution is specific in each lighting condition. For such results mentioned above, we could define a skin model based on one subject to generate a general skin model, and we utilize an elliptical boundary to fit the skin cluster on C_r-C_b subspace, which is validated in [5]. Equation (1) and (2) show the elliptical decision boundary, and Fig. 2 is the illustration of skin-tone pixels extraction.

where Cx and Cy are ellipse center(x,y); A and B are major

$$\begin{bmatrix} Cr' \\ Cb' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} Cr - Cx \\ Cb - Cy \end{bmatrix} \quad (1)$$

$$Skin(x,y) = \begin{cases} 1, & if \quad \frac{(Cr' - Cx)^2}{A^2} + \frac{(Cb' - Cy)^2}{B^2} <= 1 \\ 0, & otherwise \end{cases} \quad (2)$$

and minor axis of ellipse; and θ is rotation angle of ellipse.

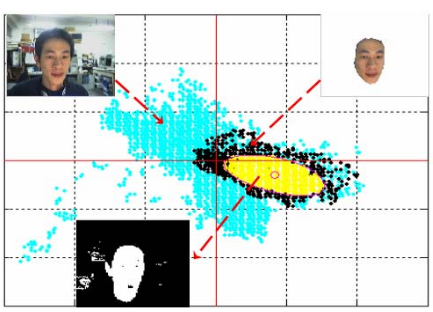

Fig. 2. A illustration of skin detection using ellipse model on C_r-C_b subspace. (The cyan dots represent the pattern of entire image, the black dots are exact face pixels, and yellow dots are defined by pre-trained ellipse model.)

B. Recognition of Illumination Conditions

Since skin-based face detection is luminance-dependent, the skin-tone distribution would be different in each illumination condition. The effects of illumination variation would dramatically decrease the stability and accuracy of skin-based face tracking system. Hence, in this study we present an effective illumination recognition technique that makes the system have an ability to derive an optimal skin model to do the face tracking task. We employ K-Nearest Neighbor (KNN) classifier for distinguishing different illuminations, and each illumination has a specific skin model to extract the skin patches in images. For this perception, we define six features in KNN to identify the surrounding illumination condition, including center of skin-tone cluster and percentages (Pi) of the skin-tone distribution in four quadrants on C_r-C_b subspace:

$$P_i = \frac{skin(C_r, C_b) \in quadrant_i}{\sum_{i=1}^{4} skin(C_r, C_b) \in quadrant_i} \times 100\% \quad (3)$$

After defining KNN features for recognizing illumination conditions, we trained an elliptical model with 10 images under per illumination condition to extract skin-tone pixels (see Fig. 3A), and the coefficients of elliptical model include the length of major and minor axis, center (x, y), and rotation angle.

Further, we use un-trained image sets, 30 images per environment, to evaluate the feasibility of KNN recognition task and quantify the efficiency of skin extraction. The experiment shows that the KNN classifier has a well capability for discriminating various illumination conditions to derive an optimal skin model to extract skin patches. Thus, the averaged accuracy of skin detection is around 92% (see Fig. 3B), which leads the success of face localization in images after region growing process. Based on the simulation results, we successfully verify the feasibility of KNN classifier and adaptive skin model, which overcomes illumination changes.

C. Face Localization

Skin-color distribution of face region on C_r-C_b domain has a clustered appearance, we can easily utilize a generative elliptical model to obtain skin extraction. However, the disadvantage would arise while the color of objects at the background is similar to skin-tone, which causes mis-detected problem. This situation increases the difficulties in identifying the biggest skin-tone patches as a face candidate. Hence, we use opening operation and region growing of morphological processing to decrease the mis-detected pixels (See Fig. 4).

Besides, when the area of skin-tone object at the background is larger than (or connected with) exact face region in images, the opening processing is inoperative. The mis-detection problem would occur in such situation. Thus, we adopt temporal information of video frames to eliminate the still skin-tone objects and retain the significant region of head

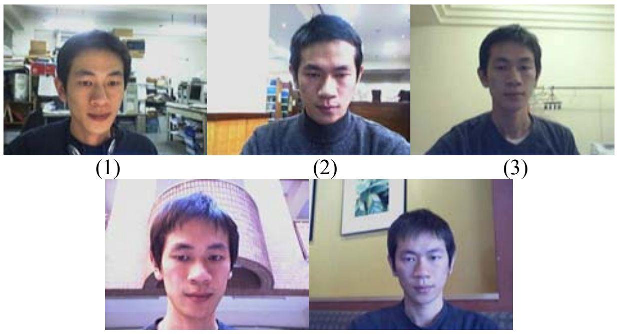
(1) (2) (3)
(4) (5)

Fig. 3-A. Five illumination environments.

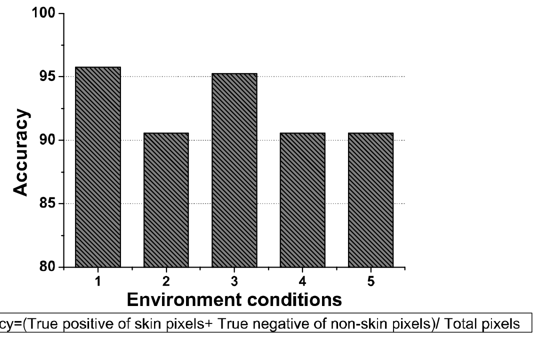

Fig. 3-B. Accuracy of skin extraction in five environments.(1) office lighting condition; (2) sunlight-inside condition; (3) darker lighting condition; (4) outdoor environment, and (5) coffee shop

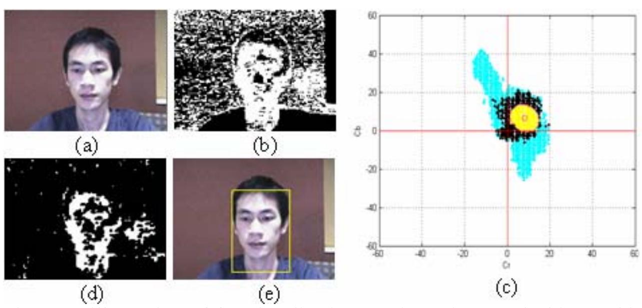

Fig. 4. A illustration of face localization. (a) input image; (b) binary skin image; (c) skin model;(d) skin image after morphological processing; and (e) face localization.

$$SDF(x,y)=\begin{cases}1, & if \sum_{i=1}^{N}\left|F_i(x,y)-F_{REF}(x,y)\right|\geq 255\\ 0, & otherwise\end{cases} \tag{4}$$

rotation movements. This technique is based on motion-based detection method utilizing sequence frames subtraction [7], as in (4).

Initially, the first frame in frames interval N is obtained to be a reference frame F_{REF} at index i. Then, from frame $F_{(i+1)}$ to frame $F_{(i+N)}$ are subtracted by F_{REF} and have a threshold at suitable value 255 to obtain a sum of difference frames variable, SDF. By this way, the still skin-tone objects at background would be eliminated at time interval. In Fig. 5, it shows the result of motion-based detection method, reinforcing the disadvantage of skin-based detection method, to obtain face tracking task.


(a) (b)

Fig. 5. Face localization without and with motion-based detection method shown in (a) and (b) respectively.


(a) (b) (c)

Fig. 6. Result of eye localization.(a) Y band image; (b) *EYEcandidate* image; (c) eye localization.

D. Eye Tracking

After identifying face region on successful captured frames, we can efficiently detect eye features based on Y component in $YCbCr$ color space. Iris usually exhibits low intensity of luminance despite different environments, and detection of sharp changes in Y component would give more stable efficiency. For this reason, we calculate mean and standard deviation according to Y component of face candidate to identify these region where gray-level intensity of inherent pixels is significant different, as in (5).

$$EYE_{Candidate}(x,y)=\begin{cases}1, & if\ I(x,y)<(Face_{Mean}-Face_{Std})\\ 0, & otherwise\end{cases} \tag{5}$$

Besides, since darker region of irises and eyebrows would be simultaneously detected, the position constrain of knowledge-based information, eye feature with lower position, is also applied to localize eye candidates. (Fig. 6)

E. Cursor Control Strategy

Based on head rotation of users, the position of eyes center would change according rotation direction, but the center of face region would roughly maintain in still position. Thus, we utilize relative motion vector between eyes center and face center to control the computer cursor via head rotation, as in (6), and it can avoid user position shifts. Besides, we set a tolerance *X Range* and *Y Range* of cursor control to decrease involuntary head vibration, which makes the mouse cursor more stable in displacement, as Fig. 7.

$$Condition(x,y)=(E_{Center}(x,y)-F_{Center}(x,y)-P_{Ref}) \tag{6}$$

where $E_{Center}(x,y)$ and $F_{Center}(x,y)$ represent the center of eyes and face respectively, and is the reference point of relative motion vector between $E_{Center}(x,y)$ and $F_{Center}(x,y)$ at previous frame. Afterwards, we define nine strategies of cursor control, and then the obtained $Condition(x,y)$ would derive the direction and displacement of cursor on the PC screen (Fig. 7).

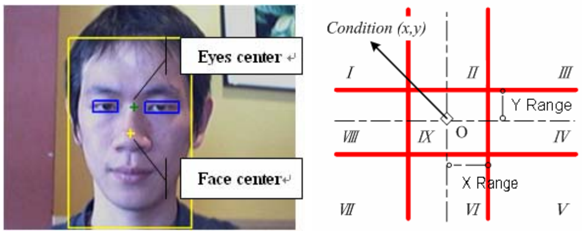

Fig. 7. Reference points and strategies of cursor control.

III. Implementation

We implement the Webcam Mouse system on laptop PC with a Pentium Ⅳ 2.4-GHz CPU and the captured device is Logitech Webcam. The captured frame format is 320x240, frame rate is 15 frames per second, and the system take up 45% of window resources.

In the results, we have successfully demonstrated that the system can track user face and eye features under various environments with complex background, such as office, external sunlight environment, darkness environment, outdoor, and coffee shop, shown in Fig. 8. As results, despite environments change, based on illumination recognition and adaptive skin models the system still can obtain the visual tracking of face and eye features. Furthermore, the ambient person would not affect the system stability because procedure of the face localization adopts the biggest connected skin-patches as user face.

IV. Discussion

In this paper we propose using KNN classifier to determine various illumination conditions. The procedure of illumination recognition would be started at time interval of tracking task rather than per sequence frame. Hence, this advantage of simplicity in algorithm allows parallel tasks on the notebook platform. The User Interface (UI) of real-time skin model generator also helps to increase the flexibility of this system. Although KNN classifier only initiates five environments at preliminary stage, the system permits users to define and add their favorite environments to KNN for computer access.

Due to current eye tracking method is based on darker iris' feature; it won't work well for dark eyeglasses at this stage. We are now adopting other techniques to overcome this issue.

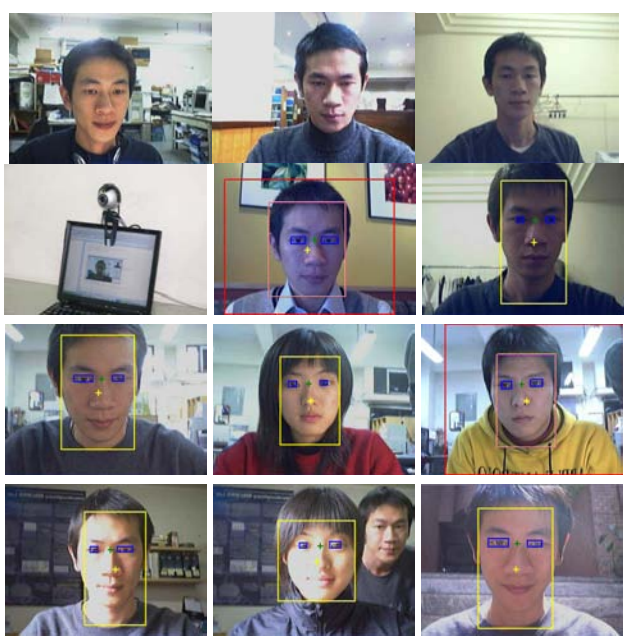
Fig. 8. Results of face and eye tracking in various environments.

V. Conclusion

In this study, we have proposed the usage of KNN classifier to determine various illumination conditions, which is more feasible than lighting compensation processing in real-time implementation. Five typical illumination environments are used as a starting point to automatically generate optimal parameters for face tracking model. We have demonstrated that the accuracy of face detection based on the KNN classifier is higher than 90% in all various illumination environments. In real-time implementation, the system successfully tracks user face and eyes features at 15 fps under standard notebook platforms.

This Webcam Mouse system is still under development, further efforts are focused on low CPU loading, high tracking accuracy and friendly usage for the disabled to provide access PC in information retrieval through internet and multimedia entertainment. Preliminary studies in "point and click" style PC web games also shows promising applications in consumer electronic markets in the future.

Reference

[1] C.Y.Chen and J.H.Chen, "A Computer Interface for the Disabled by Using Real-Time Face Recognition," *Proceedings of the IEEE 25th Annual International Conference on Engineering in Medicine and Biology Society*, vol. 2, pp. 1644-1646, 2003.

[2] D.O.Gorodnichy and G.Roth, "Nouse "Use Your Nose as a Mouse" Perceptual Vision Technology for Hands-Free Games and Interfaces," *Image and Vision Computing*, vol. 22, no. 12, pp. 931-942, 2004.

[3] M.Betke, J.Gips, and P.Fleming, "The Camera Mouse: Visual Tracking of Body Features to Provide Computer Access for People with Severe Disabilities," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 10, no. 1, pp. 1-10, 2002.

[4] M.H.Yang, D.Kriegman, and N.Ahuja, "Detecting Faces in Images: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34-58, 2002.

[5] R.L.Hsu, M.Abdel-Mottaleb, and A.K.Jain, "Face Detection in Color Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696-706, 2002.

[6] S.C.Pei and C.L.Tseng, "Robust Face Detection for Different Chromatic Illuminations," *Proceedings of IEEE International Conference on Image Processing*, vol. 2, pp. 765-768, 2002.

[7] T.Funahasahi, M.Tominaga, T.Fujiwara, and H.Koshimizu, "Hierarchical Face Tracking by Using PTZ Camera," *Proceedings of the Sixth IEEE Conference on Automatic Face and Gesture Recognition*, pp. 427-432, 2004.