JOURNAL OF KOREA MULTIMEDIA SOCIETY, VOL. 10, NO. 12, DECEMBER 2007(pp. 1591-1600)

# Human Head Mouse System Based on Facial Gesture Recognition

Li Wei[†] and Eung-Joo Lee[††]

## ABSTRACT

Camera position information from 2D face image is very important for that make the virtual 3D face model synchronize to the real face at view point, and it is also very important for any other uses such as: human computer interface (face mouth), automatic camera control etc. We present an algorithm to detect human face region and mouth, based on special color features of face and mouth in $YC_bC_r$ color space. The algorithm constructs a mouth feature image based on $C_b$ and $C_r$ values, and use pattern method to detect the mouth position. And then we use the geometrical relationship between mouth position information and face side boundary information to determine the camera position. Experimental results demonstrate the validity of the proposed algorithm and the Correct Determination Rate is accredited for applying it into practice.

**Keywords:** Face Detection, Camera Position Determination

## 1. INTRODUCTION

To use mouse and keyboard as human-computer interface have been a very long history since computer been invented. As society been developing, while camera becomes standard configuration for personal computer (PC) and computer speed becomes faster and faster, achieving human-computer interface using computer vision becomes a feasible solution. Through this can make human hand freely, and also for disability of hand people, they also can use computer optionally. For hand-free solutions, one feasible solution is to track human body movement in video input, through utilizing the body video the intention of human can be inferred by inferring algorithm, and then computer can respond to it [1,2].

Usually, there are several parts of body can be used for tracking and perceptual user interface such as: human face, human mouth, human hand, eyes, etc. there are also most kinds of algorithm for these parts. [3-6] have proposed algorithms to use movement of eyes to navigate mouse, [7,8] proposed algorithms used nose movement, and [9,10] used nostrils, etc. But for nose, it is very difficult to detect it because the feature of nose is not very clearly to face. And for eyes, when human rotated head to the profile side, one of eyes is not in the captured image and when human rotated head in z-axis, the middle point of two eyes is difficult to confirm. Similarly, the feature of nostrils is also not very obvious in face image. So to detect these organs feature is not the best method used for tracking and perceptual user interface.

But as we have observed that, human mouth contains very strongly independent feature to the

※ Corresponding Author : Eung-Joo Lee, Address : (608-711) 535 YoungDang-Dong, Nam-Gu, Busan, Korea, TEL : +82-51-610-8372, FAX : +82-51-610-8846, E-mail : ejlee@tu.ac.kr

Receipt date : Apr. 30, 2007, Approval date : Oct. 20, 2007

[†] Department of Information Communication Engineering, TongMyong University, Korea (E-mail : 12li-wei@163.com)

[††] Department of Information Communication Engineering, TongMyong University, Korea

※ This work was supported by the SKTU Institute of Next Generation Wireless Communications funded by SK telecom(SKTU-07-004).

other region in face image, even in profile side of face image, and in face region, it always have a corresponsive and static position when human rotate or shake head. And also we have observed that the track of mouth is different when human rotate head or shake head. Usually different head motion with different mouth track, for example, when human view things shown in monitor, usually human head move slowly and correspondingly for the track of mouth, the period is longer than that human shake head, and also the track is so different when human shake head to right, left, top or bottom. So the mouth track information can be used for tracking and perceptual user interface.

In this paper, we will present an algorithm which used the mouth track information to retrieve head motion parameters for mouse control. This algorithm utilizes only one camera as video input, and can retrieve the head motion parameters by capturing track of mouth, and trough analyzing the feature of track signal to separate movement information and command information (left click or right click or double click) for mouse command, for the movement information of mouse, a mapping algorithm will be used to map the mouth position in face region MMA to mouse position in the screen. And for the command information, we will employ a recognition machine based on neural network to recognize these command signals for mouse. We designed 3 different mouse control command: left click, right click and left double click. In the experiments, the controllability of the 3 mouse control command is tested in Windows XP environment.

The rest of this paper is organized as follows: The face region and mouth position detection algorithm is detailed in section 2. Mouse command information determination algorithm is detailed in section 3. And section 4 gives the tests results to justify the efficiency of the proposed algorithm. Section 5 gives conclusions.

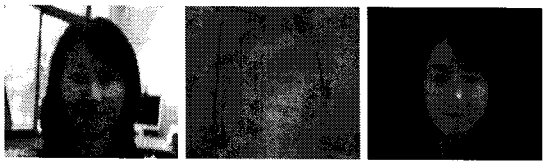
## 2. FACE REGION AND MOUTH POSITION DETECTION ALGORITHM

Firstly, the most important of our algorithm is to detect face region and mouth position from the input image of video. So in this section, an efficient and fully automatic framework is proposed for face region detection and mouth position detection from an input single human face image. The framework consists of three parts: 1) Face region detection for face candidates, 2) Mouth movement region confirm and 3) mouth position detection in the detected face region image. The following subsections will describe these three parts in details.

### 2.1 Face Region Detection

For detecting face region, [11] have introduced some algorithms, and [12] compared advantage and disadvantage of these algorithms by a table. For considering the accuracy and efficiency of algorithms, the skin color information is better than others. So we will use the skin color information to detect human face. Usually the normalized red-green (r-g) space is not the best choice for face detection because the feature of face region does not very strongly independent of the background in RGB color space, but we can find that in the $YC_bC_r$ color model, the $C_r$ values of face color are strongly independent of the background. Therefore, we can translate the input image from RGB color model to $YC_bC_r$ color model, and the $C_r$ values can be employed to detect the region of face. According to our framework, we enactment that face region possess 70% of total region, so the histogram method as shown in equation 1 can be employed to detect $C_r$ threshold of face. And Fig.1 shows the detection results.

$$S = \sum_{i=0}^{255} Hist(i), \quad Cr_{Th} = 70\% \times S \tag{1}$$



(a) (b) (c)

Fig. 1. Face Detection Result: (a) Source Image, (b) Cr Image of Face and (c) Face Region Image.

## 2.2 Mouth Movement Area (MMA) Confirm

Usually, human mouth is on below-middle of face region, and it is about 1/4 of total face height, and in horizontal direction it is in the middle of face width, so for our algorithm we can enactment that MMA is 1/4 of face region in height and whole in width. Fig. 2 shows a face region image and the MMA

## 2.3 Mouth Position Detection in Detected Face Region Image

In the selected face MMA We can notice that the $C_b$ and $C_r$ cluster of face have nearly reversed relationship in the $YC_bC_r$ color space because the color of mouth region contain stronger red component and weaker blue component than other facial region. So when face color be translated to $YC_bC_r$ color space, chroma of face color is strong and luma is weak. But in mouth region, the difference between chroma and luma is not very evidence than other regions of face. This characteristic can also be observed in $C_b$-Y and $C_r$-Y subspaces shown in Fig. 1 (f) (g) in deatils. And we further notice that the mouth has a relatively low response in the $C_b^2$ feature, but it has a high response in $C_r^2$. So in mouth region the $C_b^2$ and $C_r^2$ will be closed and on other facial region they will be nearly different. Then we can make a multiply operation between $C_b^2$ and $C_r^2$, so in the multiplied face image, mouth region contains stronger information than other face region. The mouth mapping operating equations are shown as following:

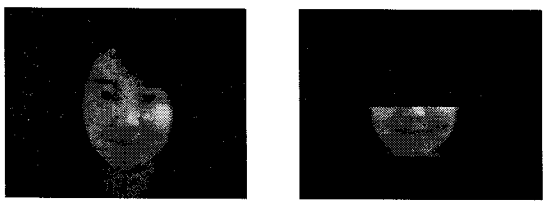
$$M_{Mouth_{Map}} = \frac{(\frac{C_r^2}{\phi_1} \cdot \frac{C_b^2}{\phi_2})}{\phi_3} = \frac{C_r^2 \cdot C_b^2}{\phi_1 \cdot \phi_2 \cdot \phi_3} \quad (2)$$

$$\phi_1 = 1.3 \cdot \frac{1}{M \cdot N} \sum_{(i,j) \in M,N} C_r, \ \phi_2 = 1.3 \cdot \frac{1}{M \cdot N} \sum_{(i,j) \in M,N} C_b,$$

$$\phi_3 = 1.3 \cdot \frac{\phi_1 + \phi_2}{2} \quad (3)$$

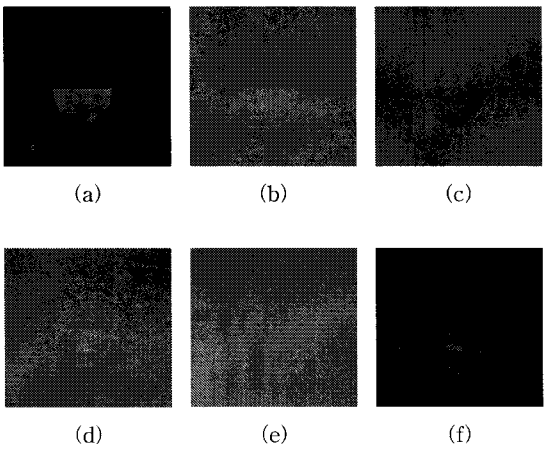
Here both $C_r^2$ and $C_b^2$ are normalized to the range [0,255], and M, N is width and height of MMA The parameter $\Phi_1$ is estimated as an average value of $C_r$ and $\Phi_2$ is an average value of $C_b$ and $\Phi_3$ is an average value of $C_r$ and $C_b$. Fig. 3 shows the operating results.

Then we can use a pattern to detect the mouth region. We select a pattern which size with 1/3 of



(a) (b)

Fig. 2. Face Region Image and MMA: (a) Face Region Image and (b) MMA.



(a) (b) (c)

(d) (e) (f)

Fig. 3. Mouth Mapping Results: (a) MMA Image, (b) Cr Image of MMA, (c) Cb Image of MMA, (d) Cr2 /Φ1 Image of MMA, (e) Cb2 / Φ2 MMA, and (f) Mouth Mapping Image.

width and 1/4 of height to detect the mouth region because as we have observed that usually mouth region hold 1/3 of total width and 1/4 of total height. The mouth pattern can be moved from left-top point to right-bottom point in Mouth Mapping Image, and then pixels value in each pattern can be integrated, the maximal integrated value can be searched. Then we can confirm that the mouth region is contained in this maximal integrated value pattern. Then we can calculate the center point position (also the mouth position) of this pattern. Fig. 4 shows the pattern operating result and these operating equations are shown as following:

$$\underset{(m,n)\in M,N}{Hist(m,n)} = \sum_{i=1}^{\frac{1}{3}\cdot M}\sum_{j=1}^{\frac{1}{4}\cdot N} Mouth_{Map}(m+i,n+j) \tag{4}$$

$$Value_{Max} = \underset{(m,n)\in M,N}{Max(Hist(m,n))} \tag{5}$$

$$P_{mouth} = Position(Value_{Max}) \tag{6}$$

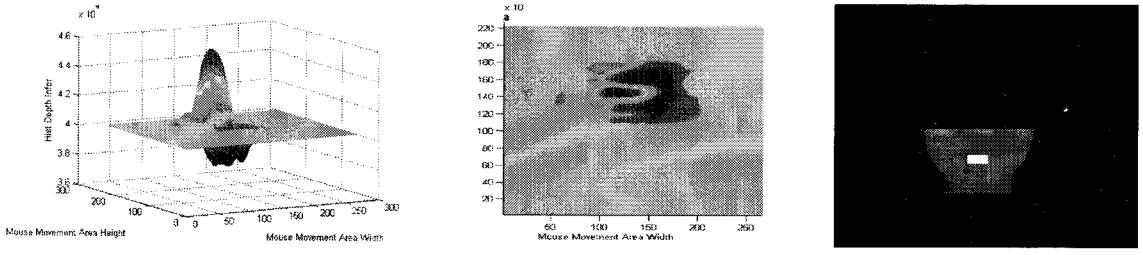
## 3. MOUSE COMMAND INFORMATION DETERMINATION FROM FACE VIDEO

Usually, the most important information of mouse is movement, left click, right click and double left click command information, so our main job is to calculate these mouse command information used the face mouth movement track in real-time video captured from camera. And the following subsections will discuss the algorithm in details: 1) mouse movement information and command information division. 2) Mouse movement information determination. 3) Mouse command information determination.

### 3.1 Mouse Movement Information and Command Information Division

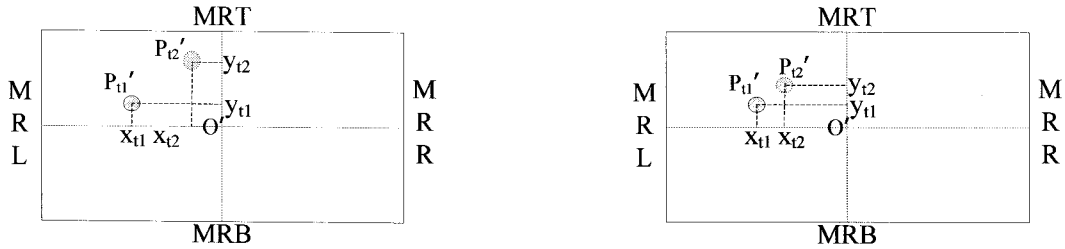
We can observe that when people see something in the screen, his head usually move slowly, so the mouth also moves slowly in the MMA, and we can assume that this is mouse movement command information. And when human nodded or shake head, we can also observe that his mouth moved quickly in the face region, we can assume that it is mouse command information. So, for the input head video, when human head move slowly, the mouth position interval from one video frame to the next video frame is also short than move quickly. For considering that mouse command information must be real-time control command, our algorithm should not waste much time to determine it. So we can use mouth interval information between front-frame and the next-frame in video frames to confirm if a mouth track is movement information or command information, as the sketch map shown in Fig. 5. The interval threshold can be specified through experiments.

The interval information in horizontal direction can be calculated by equation 7, and in vertical di-




(a) (b) (c)

Fig. 4. Pattern Operating Results for Mouth Mapping Image Shown in Fig 4(f): (a) 3D Diagram of Result, (b) Overlook of Depth Information for All Patterns (colors which changed from crimson to navy blue represent the depth information), (c) Pattern Operating Result Image.




(a) (b)

Fig. 5. Sketch Map of Interval Information between Front-frame and the Next-frame in Video Frames: (a) Interval Information for Mouse Movement Information.

rection by equation (8):

$$S_x = P'_{t2}(x) - P'_{t1}(x) \tag{7}$$

$$S_y = P'_{t2}(y) - P'_{t2}(y) \tag{8}$$

So the mouse command information can be determined by equation 9 and 10 in both horizontal and vertical direction.

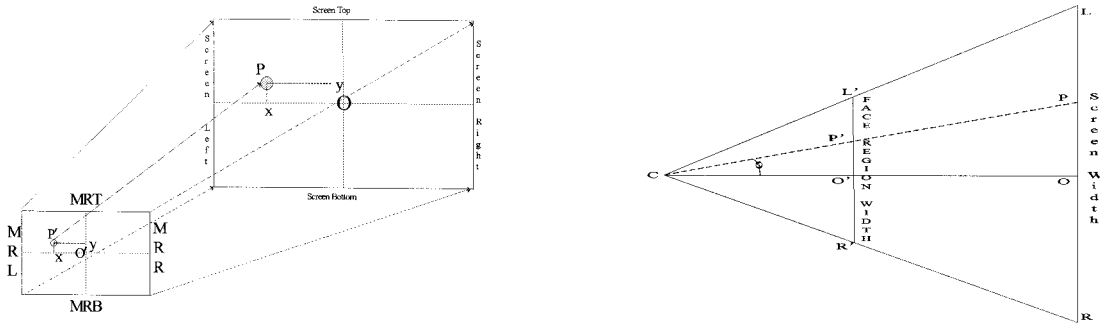
$$MouseCom_x = \begin{cases} 0, \{|S_x| < Th_x\} \\ 1, \{|S_x| > Th_x\} \end{cases} \tag{9}$$

$$MouseCom_y = \begin{cases} 0, \{|S_y| < Th_y\} \\ 1, \{|S_y| > Th_y\} \end{cases} \tag{10}$$

Here if value of $MouseCom_x$ is 0, that means it is mouse movement information in horizontal, and we can set the new point of mouse position in horizontal direction by using mapping equations (12) which we will discuss in subsection 3.2, also if $MouseCom_y$ is 0, the new point of mouse position in vertical direction can be set by mapping equations (13). And correspondingly if $MouseCom_x$ is 1, that means it is mouse click command information, and we will record the mouth movement track and use a recognition machine to recognize the mouth track and gain mouse command information. This approach will be discussed in subsection 3.3 in details.

### 3.2 Mouse Movement Information Determination

If mouth movement track is judged as mouse movement information, the next job is to map the position of mouth in MMA to screen as mouse position. So we must set up a mapping relationship between MMA and screen. As shown in Fig. 6. We can assume that in the projection region the mouth




(a) (b)

Fig. 6. Sketch Map of Mapping Relationship between MMA and Screen: (a) 3D Perspective Sketch Map. (b) 2D Plane View of Sketch Map.

position is P'(x, y), and in the corresponding screen region, the mouse position is P(x, y). So we can set up a linearly equation as following:

$$\frac{|P'O|}{|PO|}=\frac{|CO'|}{|CO|}=\frac{|L'R'|}{|LR|} \tag{11}$$

And for horizontal direction: $|P'O|=P'(x)-P'_O(x)$, and $|PO|=P(x)-P_O(x)$, so we can get an equation from 10 as shown in equation (12)

$$\frac{P'(x)-P'_O(x)}{P(x)-P_O(x)}=\frac{|L'R'|}{|LR|} \tag{12}$$

So we can get the final mapping equation for horizontal direction as following:

$$P(x)=\frac{(P'(x)-P'_O(x))\cdot|LR|}{|L'R'|}+P_O(x) \tag{13}$$

And for vertical direction is as equation (14):

$$P(y)=\frac{(P'(y)-P'_O(y))\cdot|TB|}{|T'B'|}+P_O(y) \tag{14}$$

Here |LR|, |TB| are width and height of screen which is measured by pixels. And |L'R'|, |T'B'| is width and height of MMA which is also measured by pixels. $P_O$ is center point of screen and $P'_O$ is center point of MMA. Therefore the mouse can be navigated by the position of mouth in MMA.

### 3.3 Mouth Command Information Determination

Once we have detected that a mouth track is a command Information, but there are several kinds of command information for mouse such as: left click, right click, left double click and right double click command, how to distinguish these command information is very important for mouse. In our algorithm, we defined human head shaking motion for mouse command information, and mouth trace information can be employed to distinguish these command information because when human shake head with different motion, and correspondingly, the mouth track is also so different to each others.

Firstly, we can define head shake motion for mouse command as following: 1. Left single click: head shake to left once; 2. Left double click: head shake to left twice; 3. Right single click: head shake to right once; (4) Right double click: head shake to right twice.

Then we can capture the mouth tracks when human shake head as shown in Fig. 7. The eigenvalue of mouth track information is start point, inflexion, stop point and period time, so in experiment, we can use samples library which contain these values to train the neural net, and then we can use the trained neural net to recognize practical mouth track to determine mouse command information. Here, we chose using a fast learning neural net (smart neural nets) which have been presented by to recognize the mouth track information as shown in Fig. 8.

We have developed a camera mouse system based on upper algorithm, and integrate the system with Windows XP operating system. When Windows users turn on the system, they can comfortably navigate in Windows and control mouse using their head.

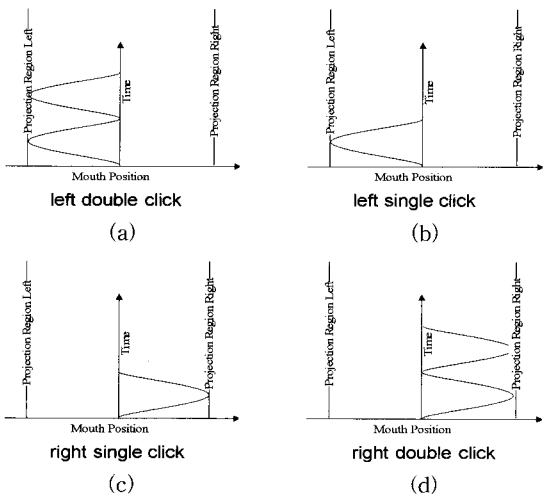



Fig. 7. Mouth Tracks Information for Mouse Command Information: (a) Mouse Left Single Click, (b) Mouse Left Double Click, (c) Mouse Right Single Click and (d) Mouse Right Double Click.

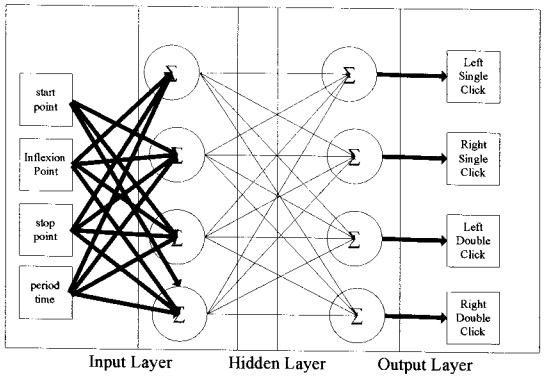
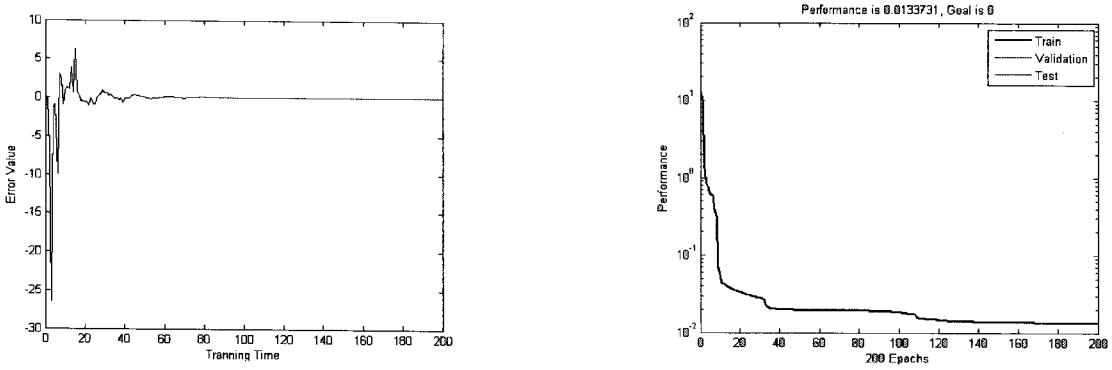



Fig. 8. Using Smart Neural Nets to Recognize Mouse Command Information.

## 4. EXPERIMENTAL RESULTS

We will implement our system which contains mouse movement control command and 4 kinds of click control command on Windows XP System to evaluate the efficient of our algorithm. We selected a screen of resolution of 1024 by 768 pixels, and camera is on the middle-top of computer monitor. As we have discussed in upper sections the most important job of our experiment is to determine the interval threshold value which is used to distinguish mouse movement command information and click command information, for our experiment, we use camera which with collection frequency of 50 Hz, so interval between two frames is 20 ms, the determination result for interval threshold value of mouse track is shown in Table 1, and for next, we will use 50 head videos for each mouse click command respectively. And totally, we will use 200 head videos to train the recognition machine based on smart neural nets technology, and set the minimal acceptance error rate is 0.001. Fig. 9 shows the training processing rate. Once the recognition machine is trained, we can use it to recognize mouth track information




(a) (b)

Fig. 9. Training Process for Mouse Click Command Information: (a) Curve of Error Rate, (b) Learning Curves of Recognition Machine based on Smart Neural Nets.

Table 1. Determination Result of Interval Threshold Value of Mouth Track (Camera with Collection Frequency of 50Hz).

| Head Move Direction | Number of Test Video for Command | Number of Test Video for Movement | Average Interval Threshold Value for Command | Average Interval Threshold Value for Movement | Average Interval Threshold |
|---|---|---|---|---|---|
| Left | 25 | 25 | 1.544 | 3.894 | 2.719 |
| Right | 25 | 25 | 1.575 | 4.052 | 2.8125 |
| Top | 25 | 25 | 0.453 | 2.478 | 1.4655 |
| Bottom | 25 | 25 | 0.367 | 2.256 | 1.3115 |

and then determine mouse click command. Here we define the recognition machine outputs meaning for mouse click command as shown in Table 2. And Table 3 gives the recognition results of 200 head motion videos for mouse click command information.

Table 2. Recognition Machine Outputs Meaning for Mouse Click Command.

| Mouse Click Command | Recognition Machine Output |
|---|---|
| Left Single Click | 1 |
| Right Single Click | 2 |
| Left Double Click | 3 |
| Right Double Click | 4 |

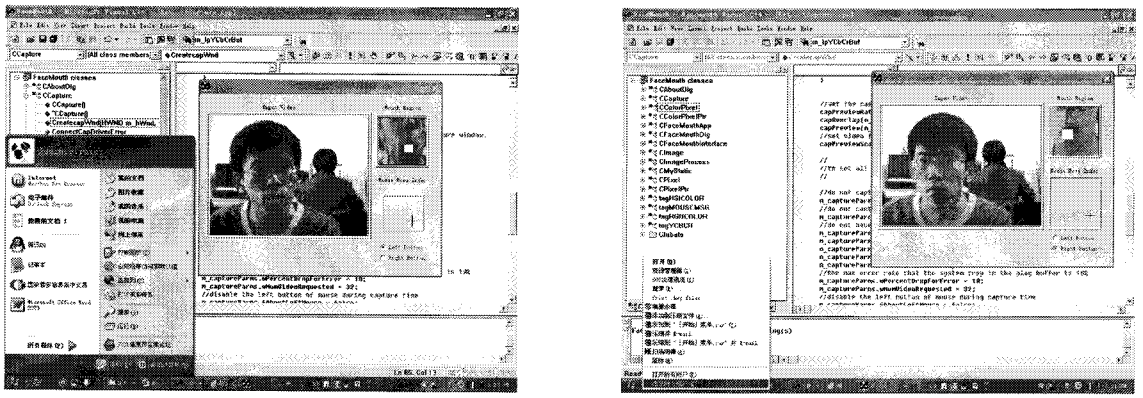
Table 3. Recognition Results for Mouse Click Command Information.

| Head Motion Video | Number | Average Result Output | Correct Rate |
|---|---|---|---|
| shake to left once | 50 | 1.023 | 97.25% |
| shake to right once | 50 | 1.994 | 96.75% |
| shake to left twice | 50 | 2.985 | 98.12% |
| shake to right twice | 50 | 4.052 | 96.24% |

Since our System is integrated to Windows XP Operation System, we can easily use camera to navigate in Windows System. Fig. 10 shows the application that using camera mouse to navigate in Windows XP system.

## 5. CONCLUSIONS

In this paper, we have presented a camera mouse technology based on mouth tracking information from a single 2D face image using the relationship between mouth position information and face region boundary information. Our system firstly nonlinearly transformed the input video frame of human head into $YC_bC_r$ color space, then use the visible chrominance feature of face in this color space to detect human face region. And then for face candidate, use the nearly reversed relationship information between $C_b$ and $C_r$ cluster of face feature to detect mouth position. Then we use mouth position interval information between front-frame and the next-frame in video frames of head motion to confirm if a mouth position track is movement information or command information for mouse, once movement information been determined, we use a mapping function which is set up by relation-



(a) (b)

Fig. 10. Application that Using Camera Mouse to Navigate in Windows XP System: (a) Using Camera Mouse to Control Mouse Cursor Move in Windows and Using Single Left Click of Camera Mouse to Open Start Menu, (b) Using Right-click of Camera Mouse to Open Resource Menu in Windows.

ship between MMA and screen to map mouth position in MMA to screen, then we can get the real mouse position information, so we can operate mouse cursor to the new position in OS. On the other ways, if mouse click command information been determined, we will capture the mouth movement track, and then use a recognition machine to recognize these track information to determine which click command the mouth movement track is. Then we have made experiments in Windows XP OS to verify the effectiveness of mouse navigation system based on presented algorithm. The experiment results show that our presented algorithm can get a perfect effect for using the camera as mouse, and human can comfortably use their head to control mouse in Windows.

Our finally object is to develop a system which can achieve all of mouse functions such as: mouse drag, mouse wheel scroll to upper and scroll to down. And an interesting future direction is to use human emotion based on the tracking of user head motions, some mouse command information can be controlled by user emotions. And also human eyes motion can be used to control mouse movement and event information.

Human head tracking information playing a very important role on camera mouse system. There are still many places in our current algorithm to be improved, such as: implementation efficiency and performance with more robust tracking in the context of camera mouse application.

## REFERENCES


[1] M. Turk, C. Hu, R. Feris, F. Lashkari, and A. Beall, "Tla based face tracking," *International Conference on Vision Interface*, pp. 229-235, Calgary, 2002.

[2] M.Turk and G. Robertson, "Perceptual user interfaces," *ACM*, Vol.43, pp. 32-34, 2000.

[3] C. Morimoto and D. Koons, A. Amir, and M. Flickner, "Realtime detection of eyes and faces," *Workshop on Perceptural User Interfaces*, San Fransisco, 1998.

[4] R. J. Jacob, "What you look at is what you get: Eye movement-based interaction techniques," *Int'l Conf. of Human Factors in Computing Systems*, pp. 11-18, 1990.

[5] P. Ekman and W. Friesen, "Facial action coding system," *Psy-chologist Press*, Palo Alto, 1978.

[6] R. Ruddarraju and etal, "Perceptual user interfaces using vision-based eye tracking," *In the 5th International Conference on Multimodal Interface*, pp. 227-233, Vancouver, 2003.

[7] D. Gorodnichy, "On importance of nose for face tracking," *In 5th Int'l Conf. on Automatic Face and Gesture Recognition*, Washington, 2002.

[8] D. Gorodnichy, S. Malik, and G. Roth, "Noose' use your nose as a mouse' - a new technology for hands-free games and interfaces," *In Int'l Conf. on Vision Interface*, pp. 354-361, Calgary, 2002.

[9] V. Chauhan and T. Morris, "Face and feature tracking for cursor control," *In 12th Scandinavian Conference on Image Analysis*, Bergen, 2001.

[10] L. El-Afifi, M. Karaki, and J. Korban, "Hand-free interface'- a fast and accurate tracking procecdure for real time human computer interaction," *In FEA Student Conference*, American University of Beirut, 2004.

[11] R.Feraud, O.J.Bernier, J.-E. Viallet, and M. Collobert, "A Fast and Accurate Face Detection Based on Neural Network," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.23, No.1, pp. 42-53, 2001.

[12] Rein-Lien Hsu and Mohamed Abdel-Mottaleb, "Face Detection in Color Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.4. No.5, 2002.


## Li Wei

Received his B. S. in Dalian University of Light Industry in China (2002~2006), and now (2006~2008) his is a M. S. student of TongMyong University in Korea. His main research interests are in image processing computer vision Biometrics and face recognition.



## Eung-Joo Lee

Received his B. S. , M. S. and Ph. D. in Electronic Engineering from Kyungpook National University, Korea, in 1990, 1992, and Aug. 1996, respectively. In March 1997, he joined the Department of Information Communication Engineering of Tongmyong University, Busan, Korea, as a professor. He has been worked as a Research Professor at Dalian Politech University in China from July 2005 to June 2006. His main research interests are in image processing, computer vision, and Biometrics. He has published many papers in the fields of Biometrics and participated in numerous government research projects.