

Moving-Object Detection Algorithm and Its Application to Video Input Device

Jui-Cheng Yen, Hun-Chen Chen, Wei-Jian Lin, and Jen-Yue Liou

Abstract—The paper proposes a moving-object detection algorithm. An auxiliary object is designed and worn on the user's brow. The user can issue commands by moving the object vertically or horizontally or opening its mouth. By segmenting the moving-object and analyzing the variation of center of gravity of the segmented object, or by segmenting user's mouth and analyzing the variation of the segmented mouth area, the algorithm can recognize up, down, left, right, and enter commands issued by the user. The software simulation results are given to demonstrate the effectiveness of the proposed algorithm. The algorithm can benefit the video input device.

Index Terms—Center of gravity, moving-object detection, segmentation, video input device.

I. INTRODUCTION

The input device [1]-[7] is an important communication bridge between user and machine. The general input devices include touchscreen, mouse, keyboard, joystick, etc. More and more friendly input devices have been developed up to now. Touchscreen is very popular and welcome because it is very easy to operate by people for all age. However, it is unfriendly to people who don't have a nimble hand. Considering this kind of need, audio [1]-[3] and video [4]-[7] approaches are two good candidates to realize input devices.

For providing vehicle control instructions, a comprehensive input device between a guidance system and an elderly driver is proposed in [2]. The audio approach is one way of issuing instructions in the device. Prof. Chen proposes a camera mouse system [4]. The system detects the user's face. Subsequently, segment the lip from the detected face and analyze the motion of the user's lip. Finally, the mouse cursor was controlled by lips movement.

In the paper, we propose a video input device. In the device, an auxiliary object is designed and worn on the user's brow. The user issues commands by moving his head vertically or horizontally or opening his mouth. The system can identify the issued commands by segmenting the moving-object and analyzing the variation of center of gravity (COG) of the segmented moving-object, or by segmenting user's mouth and analyzing the variation of the segmented mouth area (MA).

Manuscript received November 10, 2012; revised January 21, 2013. This work is supported in part by the National Science Council, R.O.C. under Grant NSC 101-2221-E-239 -038.

The authors are with the Department of Electronic Engineering, National United University, Miaoli, Taiwan, R. O. C (e-mail: {jcyen, hcchen, wjlin}@nuu.edu.tw, dt842684@yahoo.com.tw).

II. VIDEO INPUT DEVICE

To provide a friendly input device for people who don't have a nimble hand, the video input device is proposed. Lateral view of the input device is shown in Fig. 1. In the device, the video camera is placed in the front of the user and the designed auxiliary object is worn on the user's brow. Fig. 2 shows the front view of the worn auxiliary object. The color of the auxiliary object is designed to be bright red.

Five commands will be issued by opening user's mouth or moving user's head upwardly, downwardly, leftward or rightward. Since the user's head with the designed object worn on him can be regarded as a rigid body, the auxiliary object has the same movement as the head. By segmenting the auxiliary object and user's mouth, computing COG and MA, and analyzing variations of COG and MA in each acquired frame, we can correctly judge the up, down, left, right, and enter commands issued by the user.

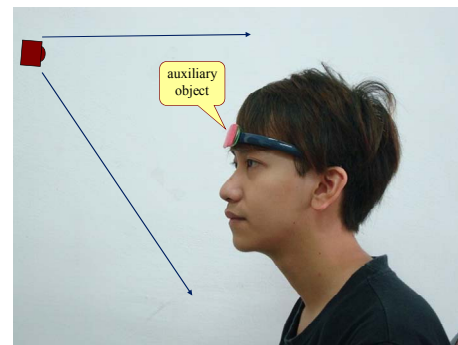


Fig. 1. Lateral view of the video input device.



Fig. 2. Front view of the designed auxiliary object worn on the user's brow.

III. MOVING-OBJECT DETECTION ALGORITHM

To perform the command decision in the video input device, the moving-object detection algorithm (MODA) is proposed.

The MODA is composed of three modes. They are (i) mode for searching equilibrium state, (ii) mode for computing reference COG or MA, and (iii) mode for

deciding issued command. Each acquired input frame is processed to segment the auxiliary object and mouth and compute center of gravity of the segmented auxiliary-object, and the segmented mouth area. Afterwards MODA enters the mode for searching equilibrium state as shown in Fig. 3. Finally, whether up/down/left/right/enter command is issued will be decided in the mode for deciding issued command.

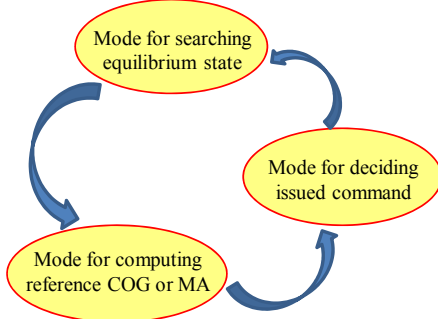


Fig. 3. Mode transition diagram in MODA.

To describe the MODA quantitatively, the following notations are adopted:

- $f_{p,q}^R$: the red component of pixel (p,q) ,
- $f_{p,q}^G$: the green component of pixel (p,q) ,
- $f_{p,q}^B$: the blue component of pixel (p,q) ,
- r_cog_i : the row-coordinate of COG of frame i ,
- c_cog_i : the column-coordinate of COG of frame i ,
- ref_r_cog : the row-coordinate of reference COG,
- ref_c_cog : the column-coordinate of reference COG,
- ma_i : the mouth area of frame i ,
- ref_ma : the reference mouth area,
- $nr_{p,q}$: the normalized red component of pixel (p,q) ,
- $ng_{p,q}$: the normalized green component of pixel (p,q) .

Based on the notations, MODA is described in the following.

A. Auxiliary Object Segmentation

Each acquired input frame undergoes auxiliary object segmentation. The auxiliary object is segmented by the following criteria:

$$f_{p,q}^R - f_{p,q}^G > 60 \text{ and } f_{p,q}^B - f_{p,q}^G > -10 \quad (1)$$

$$\text{or } f_{p,q}^R = 255 \text{ and } f_{p,q}^B - f_{p,q}^G > 0 \quad (2)$$

The second criterion is used to overcome the problem of light reflection. After the auxiliary object of frame i is segmented, the row-coordinate and column-coordinate of center of gravity, r_cog_i and c_cog_i , of the segmented auxiliary object are computed.

B. Mouth Segmentation

We modified the method proposed M. Soriano [8] to segment the user's mouth. First, the RGB color space is transformed to NCC color space [9] by the following equations:

$$nr_{p,q} = \frac{f_{p,q}^R}{f_{p,q}^R + f_{p,q}^G + f_{p,q}^B} \quad (3)$$

$$ng_{p,q} = \frac{f_{p,q}^G}{f_{p,q}^R + f_{p,q}^G + f_{p,q}^B} \quad (4)$$

Then, the following two quantities are computed

$$w(nr_{p,q}, ng_{p,q}) = (nr_{p,q} - 0.33)^2 + (ng_{p,q} - 0.33)^2, \quad (5)$$

$$Q(nr_{p,q}) = -0.776nr_{p,q}^2 + 0.560nr_{p,q} + 0.18. \quad (6)$$

Finally, the mouth is segmented by the following criteria:

$$ng_{p,q} \leq Q(nr_{p,q}), \quad (7)$$

$$w(nr_{p,q}, ng_{p,q}) > 0.001, \quad (8)$$

$$\text{and } f_{p,q}^R - f_{p,q}^G \geq 15. \quad (9)$$

After the mouth of frame i is segmented, its mouth area ma_i , the total pixel of the segmented mouth, can be computed.

C. Mode for Searching Equilibrium State

For frame i , we compute and determine whether both conditions

$$|r_cog_i - r_cog_{i-1}| \leq 3 \quad (10)$$

$$\text{and } |c_cog_i - c_cog_{i-1}| \leq 3 \quad (11)$$

Are true. If both conditions are true for three successive frames $(i-2)$, $(i-1)$ and i , it means the user's head doesn't move again and MODA enters equilibrium state. When the equilibrium state occurs at the first time, the reference COG and MA is computed from the subsequent 5 frames. Subsequently, just the reference COG is computed after the equilibrium state occurs.

D. Mode for Computing Reference COG or MA

If the $(i-2)$ th, $(i-1)$ th and i th frames are determined to be in equilibrium state, the reference COG is computed by

$$ref_r_cog = \frac{1}{5} \sum_{i+1}^{i+5} r_cog_i, \quad (12)$$

$$\text{and } ref_c_cog = \frac{1}{5} \sum_{i+1}^{i+5} c_cog_i. \quad (13)$$

The reference COG is updated adaptively for each time that equilibrium state occurs. It can overcome the problem that the user's head may be in different position at each equilibrium state.

When the equilibrium state occurs at the first time, the reference MA is computed by

$$ref_ma = \frac{1}{5} \sum_{i+1}^{i+5} ma_i. \quad (14)$$

To compute reference MA is to overcome the problem that different users of the system may have large variation in the segmented closed-mouth area.

E. Mode for Deciding Issued Command

After reference COG and MA are obtained, MODA monitors the following conditions for each acquired frame i :

$$\text{Condition 1: } r_cog_i - ref_r_cog > 50 \quad (15)$$

$$\text{Condition 2: } r_cog_i - ref_r_cog < -20 \quad (16)$$

$$\text{Condition 3: } c_cog_i - ref_c_cog > 50 \quad (17)$$

$$\text{Condition 4: } c_cog_i - ref_c_cog < -50 \quad (18)$$

$$\text{Condition 5: } ma_i - ref_ma > 200 \quad (19)$$

Conditions 1, 2, 3, 4 and 5 represent that the down, up, left right, and enter commands are issued respectively.

IV. SOFTWARE SIMULATION RESULT

To demonstrate the effectiveness of the proposed MODA, software simulations are performed. In the simulations, we record a video in our laboratory. Then, we extract 523 frames from the video by “Free Video to JPG Converter”. Each extracted frame is of size 480×640.

According to MODA, we segment the auxiliary object by the segmentation criteria for each acquired frame. As a representative, the segmentation result of the 315th frame shown in Fig. 2 is shown in Fig. 4. After the auxiliary object is segmented, the COG of the segmented object is computed. The trajectory of center of gravity of all the 523 frames is shown in Fig. 5. The segmented mouth area of all the 523 frames is shown in Fig. 6.

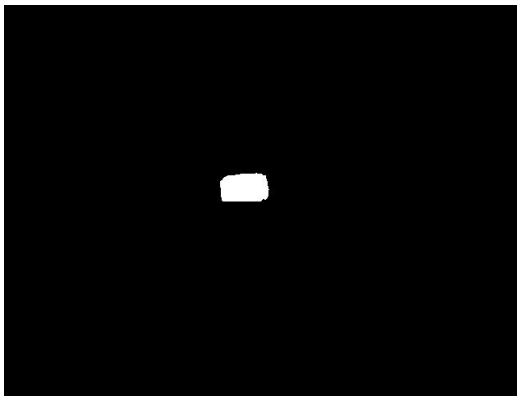


Fig. 4. Segmented auxiliary object of the twelfth frame.

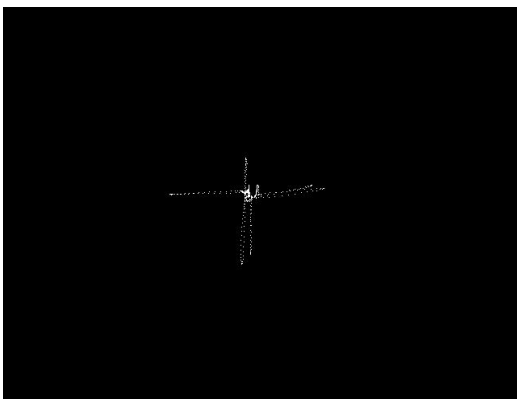


Fig. 5. Trajectory of center of gravity of the 523 frames.

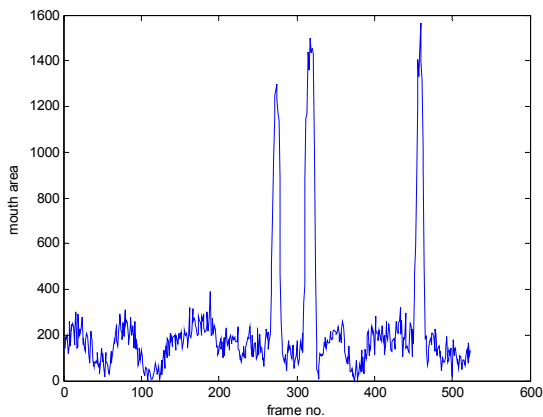


Fig. 6. Mouth areas of the 523 frames.

After performing segmentation and computing its center of gravity, each frame undergoes the examination whether the frame can be classified into the three modes according

MODA. Moreover, the command type is also determined in the mode for deciding issued command by Conditions (15)-(19). All the mode classification and command determination results of the 523 frames are shown in Fig. 7. The decided issued command is also designated in the figure by symbols “U”, “D”, “L”, “R” and “E”. Notice that symbols “U”, “D”, “L”, “R” and “E” denote the up, down, left, right and enter commands issued by the user respectively.

The complete simulation results of the 523 frames are concatenated to a video by “Window MovieMaker”. The video is uploaded to YouTube [10]. In the video, the decided issued command is represented by a red square in the user’s up, down, left, right, or upper-left position of the frame. As a representative, the simulation results of the 54th, 113th, 177th, 229th, and 274th frames are shown in Fig. 8.

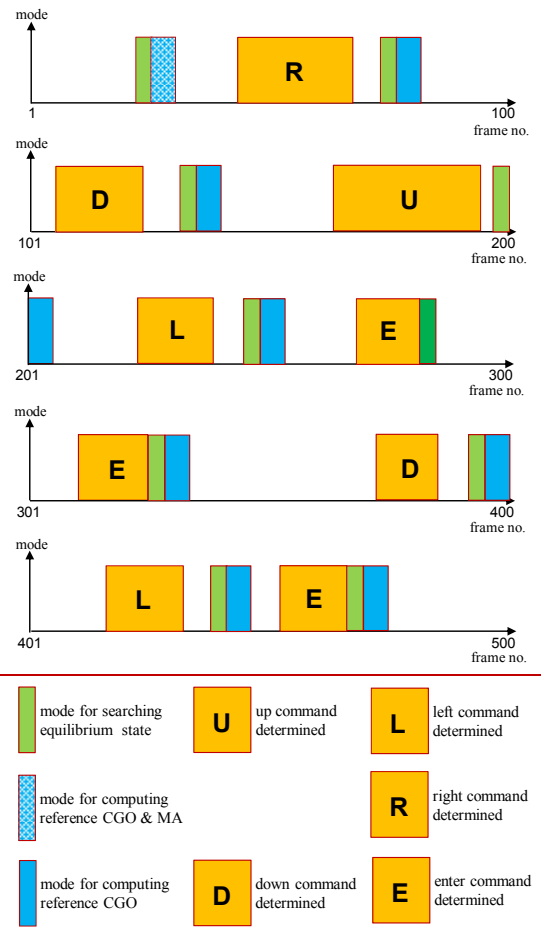


Fig. 7. All the mode classification and command determination results of the 523 frames.



(a)

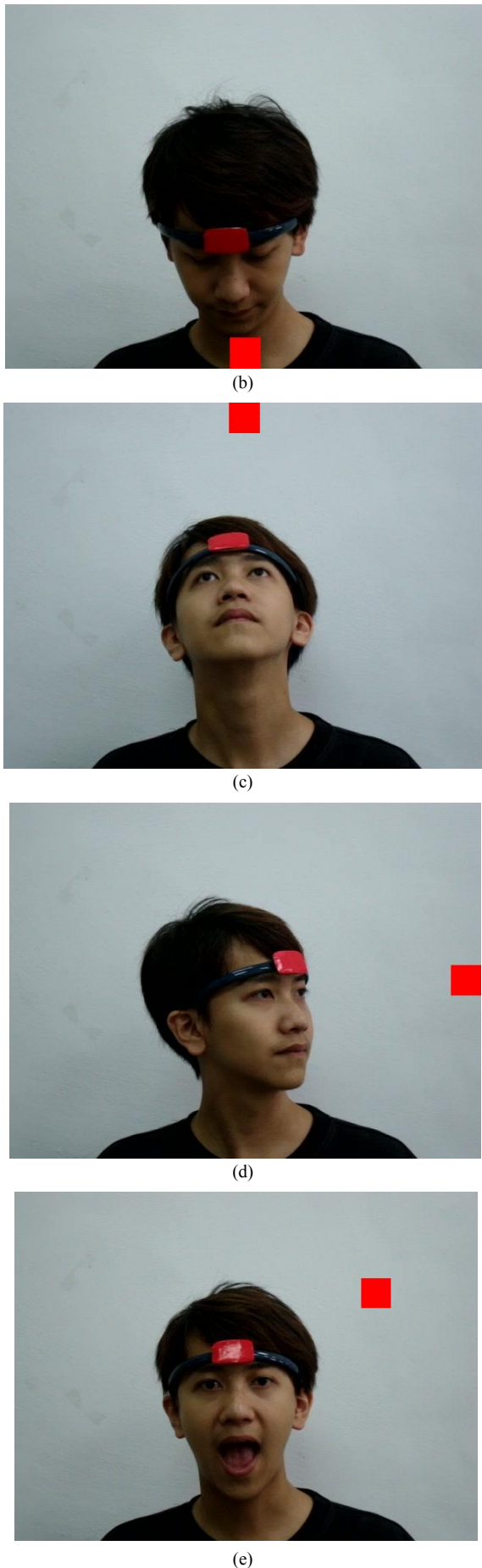


Fig. 8. (a) right command issued in the 54th frame, (b) down command issued in the 113th frame, (c) up command issued in the 177th frame, (d) left command issued in the 229th frame, (e) enter command issued in the 274th frame.

From the segmentation result shown in Fig. 4, the moving-object is well segmented. Based on the well segmentation, its center of gravity can be obtained exactly. Since the user's head with the designed object worn on his brow can be regarded as a rigid body, the trajectory of all the 523 frames' COG shown in Fig. 5 fully demonstrates the movement of the user's head. Fig. 6 shows the mouth areas of all the 523 frames. When the user issues a command by opening his mouth, the mouth region is well segmented and its area has large variation. The curve clearly indicates the user use mouth to issue commands three times and the three commands can be easily detected.

In the recorded video, nine commands are issued. Both the video shown in YouTube [10] and Fig. 7 demonstrate that all the issued commands are detected exactly. Besides, Fig. 7 also indicates the reference COG is adaptively updated in each round of command detection. Adaptive update on reference COG can overcome the problem that the user's head may be in different position at each equilibrium state. The computation of reference MA when the first equilibrium state occurs can overcome the problem that different user's lip color results in large variation of the segmented mouth area. The five representative frames shown in Fig. 8 also demonstrate the successful command detection.

V. CONCLUSION

In the paper, moving-object detection algorithm has been proposed. By segmenting the auxiliary object and user's mouth, computing COG and MA, and analyzing variations of COG and MA, the algorithm can detect the five kinds of different issued command exactly. The proposed algorithm can overcome the problems that the user's head may be in different position at each equilibrium state and different users may have large variations of segmented mouth area. Software simulation results demonstrate that all the nine issued commands in the recorded video have been detected exactly. We believe the video input device can benefit from adopting the proposed algorithm and benefit people who don't have a nimble hand.

REFERENCES

- [1] M. Pleva, S. Ondas, J. Juhar, A. Cizmar, J. Papaj, and L. Dobos, "Speech and mobile technologies for cognitive communication and information systems," in *Proc. 2nd International Conference on Cognitive Infocommunications*, 7-9 July 2011, pp. 1-5.
- [2] N. Hashimoto, S. Kato, and S. Tsugawa, "HMI for elderly drivers in an off-vehicle parking assistance system," *2008 IEEE Intelligent Vehicles Symposium*, pp. 482-487, 4-6 June 2008.
- [3] Apple. [Online]. Available: <http://www.apple.com/iphone/features/siri.html>
- [4] J. H. Chen and W. K. Lin, "A DSP based camera mouse system," M.S. thesis, Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan, 2004.
- [5] N. Conci, P. Ceresato, and F. G. B. D. Natale, "Natural human-machine interface using an interactive virtual blackboard," in *Proc. 2007 IEEE International Conference on Image Processing (ICIP 2007)*, Sept. 16 - Oct. 19 2007, vol. 5, pp. V-181 - V-184.
- [6] Y. Shi, P. Raniga, and I. Mohamed, "A smart camera for multimodal human computer interaction," in *Proc. IEEE Tenth Inter. Symposium on Consumer Electronics*, 2006, pp. 1-6.
- [7] Y. Zhang, Q. J. Liu, Y. H. Li, and L. Zeng, "Intelligent wheelchair multi-modal human-machine interfaces in lip contour extraction based on PMM," in *Proc. IEEE International Conference on Robotics and Biomimetics (ROBIO'09)*, 19-23 Dec. 2009, pp. 2108-2113.

- [8] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen, "Using the skin locus to cope with changing illumination conditions in color-based face tracking," in *Proc. IEEE Nordic Signal Processing Symposium*, Kolmarden, Sweden, 2000, pp. 383-386.
- [9] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen, "Skin detection in video under changing illumination conditions," in *Proc. 15th Inter. Conf. on Pattern Recognition*, 2000, Vol. 1, pp. 839-842.
- [10] <http://www.youtube.com/watch?v=JkZuq4nD5Fg&feature=youtu.be>



Jui-Cheng Yen was born in Tainan, Taiwan, in 1963. He received the B.S. and Ph.D. degrees in electrical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 1987 and 1992, respectively.

He is currently a Professor of the Department of Electronic Engineering, National United University, Miaoli, Taiwan. Dr. Yen was the recipient of Research Awards of National Science Council in

1996 and 1998..

His research interests include multimedia signal processing, car electronics, and VLSI design



Hun-Chen Chen was born in Taipei, Taiwan in 1961. He received the B.S. degree in electronics engineering from National Taiwan Technology University, Taipei, Taiwan, in 1990, and the M.S. and Ph. D. degrees from National Chiao-Tung University from Hsinchu, Taiwan, in 1998 and 2006, respectively.

He is currently an associate professor of the Department of Electronic Engineering, National

United University, Miaoli, Taiwan. His research interests include image and video processing, neural network application, and VLSI design.



Wei-Jian Lin was born in Taichung, Taiwan. He received his B.S. and M.S. degrees in electronic engineering from the National Taiwan University of Science and Technology, Taiwan, in 1986 and 1988, respectively, and his Ph.D. degree in communication engineering from National Chiao Tung University, Taiwan, in Jan. 2005. Currently, he is an Associate Professor at the Department of Electronic Engineering, National United University. His research

interests include the VLSI design, digital signal processing and digital communications.



Jen-Yue Liou was born in Hsinchu, Taiwan, in 1988. He received his B.S. degree in electronic engineering from the National United University, Miaoli, Taiwan, in 2011, where he is currently working on his M.S. degree. His research interests include digital image and video processing and FPGA design.