Journal of Software Engineering

ISSN 1819-4311

Academic Journals Inc.

www.academicjournals.com

Journal of Software Engineering 9 (3): 621-630, 2015
ISSN 1819-4311 / DOI: 10.3923/jse.2015.621.630

Research on the Tongue Controlled Mouse System Based on Computer Vision

[1,2]Wei Wang, [1]Wenhui Li, [2]Yongkui Liu and [2]Yanmei Xu

[1]College of Computer Science and Technology, Jilin University, Changchun, 130500, China
[2]College of Computer Science and Technology, Dalian Nationalities University, Dalian, 116600, China

Corresponding Author: Wenhui Li, College of Computer Science and Technology, Jilin University, Changchun, 130500, China

ABSTRACT

To solve the problem that people without hands can not operate computer mouse, a novel tongue controlled mouse system is proposed based on computer vision and human-computer interaction. In the system, the recognized tongue action is used to activate the mouse operation. Tongue controlled mouse system architecture is firstly introduced and then the algorithms including mouth localization, feature extraction and tongue action recognition are described. An advanced LBP operator is presented for mouth feature extraction and the SVM classifier is trained for action recognition. The experimental results show that the scheme proposed in this study can realize the tongue action recognition effectively and meet the real-time and robustness requirements of the video processing system.

Key words: Tongue controlled mouse, LBP, SVM, tongue action detection

INTRODUCTION

With the spread of computer application scope, more and more people combine their work, study and livelihood with computer directly or indirectly. But the disables who lost their hands can not operate a computer mouse and have difficulty in using a computer.

The current study of the mouse application for the disabled mainly includes the "Head controlled mouse" and "Eye controlled mouse". In 2009, the Korean HANGKYONG industry-university-institute cooperation agency applied the patent of "Head controlled mouse" (Hankyong Industry Academic Cooperation Center, 2009), using the mapping relationship of head moving and screen to control the movement of the mouse. While in April 2013, the Chongqing Institute of Science and Technology (Chongqing Academy of Science and Technology, 2013) applied "Method and system of an eye controlled mouse" patent to control mouse by using the mapping relationship of sight and screen. Besides, the somatosensory equipment Leap Motion published in 2013 uses gesture to replace mouse operation. However, through experience and comparison, these current mouse systems are proved to be not convenient or flexible enough and frequent movements of eye, finger or head will lower the experience satisfaction of system users.

This study presents a novel tongue controlled mouse system based on computer vision (Poppe, 2010) to achieve the interactive mouse application and solve the problem that people who lost their hands can not operate computer mouse. It realizes the mouse movement by moving face and sticking out the tongue to left or right to activate the mouse operation of left button click and right button click. There is no related research precedent. In this study, a new ALBP operator was

also proposed to exact the tongue action feature. Using the new operator, the tongue action can be detected effectively and the corresponding mouse action can be activated.

MATERIALS AND METHODS

System architecture: The tongue controlled mouse system architecture proposed in this study as shown in Fig. 1 includes video acquisition, face detection and tracking, the mouth area localization and tongue action recognition.

Video image acquisition: In the tongue controlled mouse system, the images for face detection and action recognition are captured from the computer user's real time motion video and the real-time video is obtained through the video capture device like camera.

Face detection: Face detection means that for any image captured by the video, the search strategy is used to identify whether it contains human faces (Wang and Wang, 2012). The Adaboost algorithm is used in the detection scheme. If a face is detected, the face rectangular area is calibrated and if the face image larger than 100×100, the mouse system starts to work. Otherwise, the system will get the next frame and keep on detecting the face. If multi faces are detected, the biggest face will be chosen as the detection area by the system.

Face tracking: Tracking is mainly to solve the matching problems of the successive image frames which are based on characteristics of the position, speed and shape and it is a continuous detection of the trajectory of the partition target and the changes in contour (Schreiber, 2008). In the process of face tracking, the center of face rectangular area corresponds to the current position of the mouse and the shifting of face moving maps to the mouse's movement. The mouse movement's mapping step can be adjusted by customer. The optical flow algorithm is adopted for face tracking in this study.

Mouth region localization and tongue action recognition: The region of the mouth is firstly positioned and then the feature of tongue action is extracted to recognize the corresponding action type.

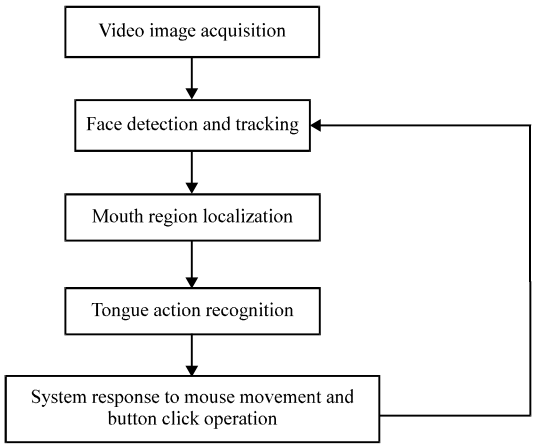

Fig. 1: Algorithm flowchart

Tongue action detection algorithm: Mouth is an important part in all the feature organs of face and mouth action recognition has vital applications in many fields, such as the human-machine interface (Chen *et al*., 2009), lip-reading and facial expression recognition and so on. This section focuses on the localization scheme of mouth region and the tongue action recognition algorithm in the human face rectangle region obtained in the face detection part.

Pose correction for face image: The tilt angle of captured human face varied widely in the video surveillance system. The pose correction technology for face image refers to process the face image captured in the video sequence to the front view. Currently there are only a few pose correction methods based on the 2D view. In order to improve the real-time ability of the algorithm, the structural information of face is used for face correction (Mao, 2008). First of all, the detected face image is identified in the video and then the position of the eyes in the frame of face image is determined and the deviated angle of the horizontal direction based on the eye position is calculated for the corresponding correction. Finally, the ratio of the corrected face area occupied in the frame to the corrected area of the face frame is calculated to determine whether the ratio is within the presented range to ensure the captured face image is always in frontage and not skew. Otherwise, the current frame should be discarded and then the system moves to the next frame.

Mouth region localization and preprocess

Mouth region localization: The positioning scheme of interested part of mouth is based on the prior knowledge (Zhang and Shen, 2008) of face structural distribution. In our scheme, the width of the mouth region is three-fifths of the width of the face framework and the mouth region is the lower one-third part of the face template detected in the face detection part.

Normalization of the mouth region image: The size of the mouth area is normalized to 32×16 which means the height of mouth image is 32 pixels and width is 16 pixels.

Grayscale processing: It refers to gray the mouth region image to obtain the corresponding grayscale image.

Illumination compensation: Illumination changes affect strongly on the action recognition rate. Mouth shading caused by uneven illumination or excessive illumination over the image will result in bigger error of identification. The Histogram Equalization Method is used for lighting pretreatment.

Characteristics of the mouth: Due to individual differences, people have different lips and different ways of sticking out the tongue, even the same individuality will have different ways at different times. The intra-class and inter-class of different tongue movements have large differences and as shown in Fig. 2, are the untreated mouth area images. Therefore, the detection of tongue action is extremely difficult.

Tongue action recognition

Feature extraction: Feature extraction algorithm is the key technology to identify the action. The Advanced Local Binary Pattern (ALBP) algorithm is proposed for feature extraction. The LBP is

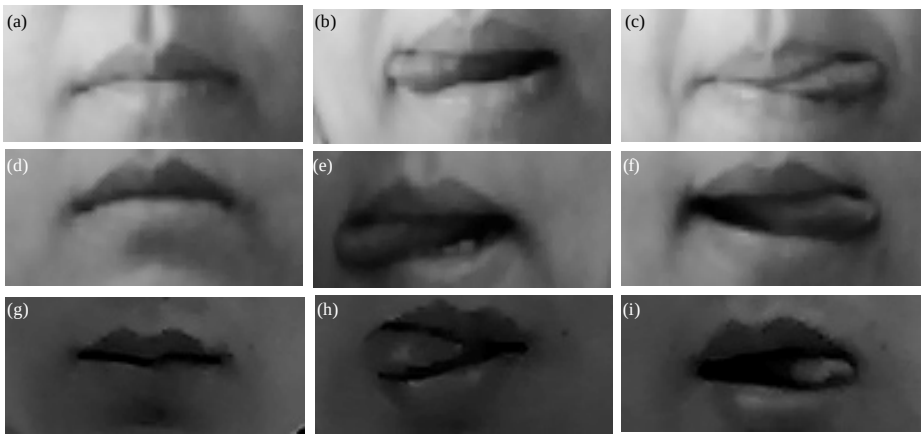

Fig. 2(a-i): Tongue action images performed by different operators, (a) Closing mouth, (b) Sticking out tongue to right, (c) Sticking out tongue to left, (d) closing mouth performed by the same operator, (e) sticking out tongue to right, (f) sticking out tongue to left, (g) Closing mouth performed by another operator, (h) Sticking out tongue to right and (i) Sticking out tongue to left

an operator used to describe the local texture features of image (Ojala *et al.*, 1996), with simple calculation, unchanged gray monotony and translation invariant. Tan and Triggs (2010) proposed an extension of the LBP texture description operator LTP (Local Ternary Pattern) which modified the threshold function of LBP operator. To some extent, it improved the noise-sensitive problem of LBP operator and achieved feature extraction under complex lighting conditions. Regional Directional Weighted Local Binary Pattern is proposed by Wang *et al.* (2011) and it has improved the accuracy of face recognition. For tongue action recognition, by viewing hundreds of mouth area images, it was found that gray scale information has a little difference between lips and tongue. Information changes are small in the horizontal direction while vertical direction information changes can reflect changes in texture and lip movements well, therefore, on the basis of paper (Wang *et al.*, 2011) the ALBP algorithm is proposed to retain more information in the vertical direction by changing the threshold function in the region of LBP operator.

The original LBP operator is defined in 3×3 windows and the windows are composed by the center pixel f_c and its eight neighboring pixels f_0,..., f_7. The relationship between them is shown in Eq. 1, in which i = 0...7.

$$G(f_i - f_c) = \begin{cases} 1, & f_i - f_c \geq 0 \\ 0, & f_i - f_c < 0 \end{cases} \tag{1}$$

And then the different weight coefficients are given to each neighboring points to get LBP code value as shown in Eq. 2:

$$LBP = \sum_{i=0}^{p-1} G(f_i - f_c)2^i \tag{2}$$

A basic LBP operator calculation process is shown in Fig. 3, in which the value of LBP is $(00100101)_2 = 37$.

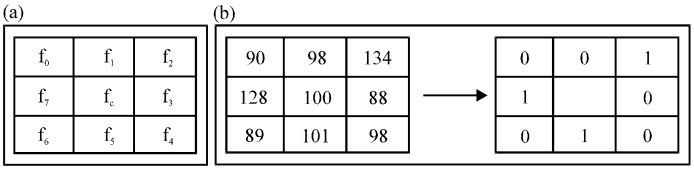

Fig. 3(a-b): Calculation method of basic LBP operator, (a) Element position and (b) Calculation of LBP

ALBP operator presented in this paper is shown in Eq. 3 below and the values of i are shown in Fig. 3a:

$$\begin{aligned} &ALBP = \sum_{i=0}^{p-1} G(f_i - f_c)2^i \\ &G(f_i - f_c) = \begin{cases} 1, & f_i - f_c \geq 2 \\ 0, & f_i - f_c < 2 \end{cases}, \quad i = 3, 7 \\ &G(f_i - f_c) = \begin{cases} 1, & f_i - f_c \geq 1 \\ 0, & f_i - f_c < 1 \end{cases}, \quad i = 0, 1, 2, 4, 5, 6 \end{aligned} \tag{3}$$

The mouth area of an image is divided into 2×2 sub-regions in this study and the experimental result shows that this division method gets the best action recognition rate. The ALBP feature of each pixel was extracted in the sub-region and operator value was used to reflect the texture region and then, establish LBP histogram feature within each sub-region. Thus a histogram can be used to describe each sub-area and the entire image feature is composed by a number of histograms and marked as LBP-Feat.

Action recognition: After using the ALBP algorithm to extract the feature, histogram feature vector can be obtained. The common way of histogram similarity measurement is the chi-square distance method, however, due to the chi-square distance algorithm requires a standard template feature vector and the tongue action varies from person to person, there is no reference to standard template. So, the SVM classifier is used to classify the action.

The Support Vector Machine (SVM) is firstly proposed by Vapnik Tang *et al.* (2011) and can be used for pattern classification and nonlinear regression. The main idea of SVM is to build a classification hyper plane as the decision surface, making the positive and negative examples isolation between the edges maximized. The SVM is a method which is good at small samples identification and generally used for smaller class classification problem.

Training of SVM classifier: In the study the SVM classifier is used for classification of three tongue actions which are closing mouth, sticking out tongue to right and sticking out tongue to left and they represent mouse movement, mouse left button click and right button click operation, respectively in the mouse system. The experiment collects 600 images and three actions including closing mouth, sticking out tongue to left and right 200 images for each. These 600 images are collected from 10 experimenters with 20 images per experimenter per action, containing images collected in different lighting sources like the daytime indoor natural light, the daytime indoor

daylight lamp and the night indoor daylight lamp, etc. Every type of image randomly chooses three quarters as the training set while the other quarter as the testing set, the SVM classifier is used for training. Three types of images were labeled as 1, 2 and 3 for classification.

In the algorithm, firstly each mouth region image is pretreated and normalized to 32×16 according to the scheme of mouth region localization and preprocess, then the tongue action eigenvector LBP-Feat is exacted by the ALBP operator from each image and eigenvector matrix of 1024×600 is obtained, finally the SVM classifier is trained for behavior classification. The experiment selects radial basis function as the kernel function of SVM. Through the experiment and parameter optimization, the classifier recognition rate reaches the highest value when the penalty factor C is set to 500 while the radial basis function parameter γ is set to 0.2.

RESULT AND DISCUSSION

Experimental results: The experiments were running in Visual C++ on Windows 7 operating system with a 2.67 GHz Intel (R) Core (TM) i5 processor and 4 GB memory. The video collection device is Sony Visual Communication camera with 0.3 mega pixels which is embedded in the notebook computer. The size of images captured by the video is 640×480 pixels.

In the experiment, as shown in Fig. 1, the system firstly starts the camera to collect video and then begins to detect face using the Adaboost algorithm and track it in the video using the optical flow algorithm, as shown in Fig. 4a, the yellow rectangle in the figure marks face bounding rectangle. And then the mouth image is detected based on the tongue action detection algorithm while the white rectangle presents the detection result of mouth image. If no face is detected, as Fig. 4b shows, the function of the system will be suspended. The mouth images detected by real-time image acquisition are shown in Fig. 5 and the three kinds of different tongue action images are closing mouth, sticking out tongue to right and sticking out tongue to left.

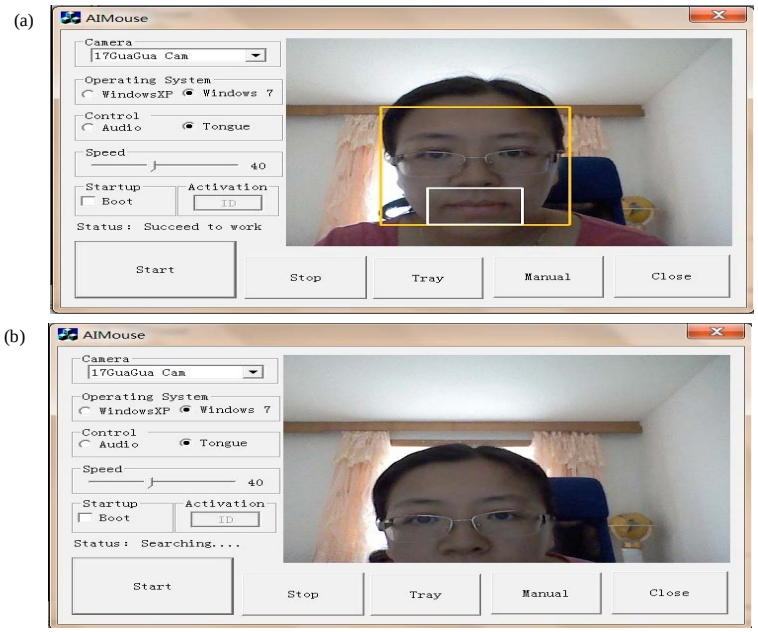

Fig. 4(a-b): Results of face detection, (a) Face was detected and then the mouse system starts to work and (b) System is suspended because no face is detected

Fig. 5(a-c): Three kinds of tongue action images, (a) Closing mouth, (b) Sticking out tongue to right and (c) Sticking out tongue to left

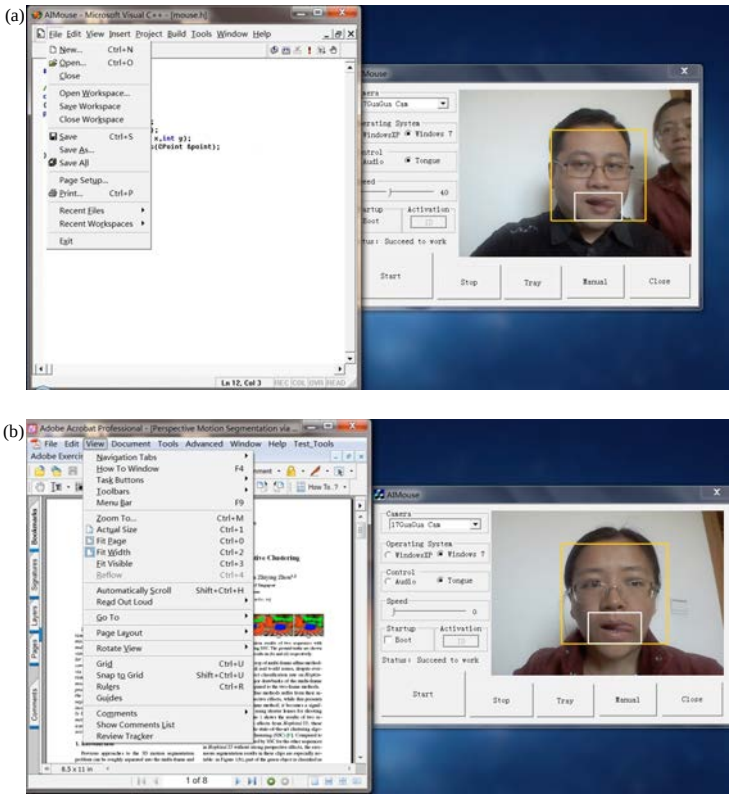

Fig. 6(a-b): Results performed by different operators

The system preprocesses and normalizes the real-time image of the mouth and uses the ALBP operator to extract eigenvector LBP-Feat of the image, then sends the eigenvector to the well trained classifier to identify the action. The output 1, 2 or 3 of SVM classifier, respectively activates the corresponding mouse behavior, namely the mouse movement, the right-click and left-click.

Figure 6, 7 are system result images under different environments, in which Fig. 6a shows the experiment result in test video 1 when user sticks out his tongue to the left and then the left mouse button is activated in its current position and the menu item "File" can be opened by clicking the button. Figure 6b shows the result in test video 2 when user sticks out her tongue to the left and open the "View" command of Adobe Acrobat. Figure 7 is the result images of the same operator in video 2 sticking out her tongue in different backgrounds and then clicking the right mouse button. As seen from the figures, the system runs well in different environments with strong robustness.

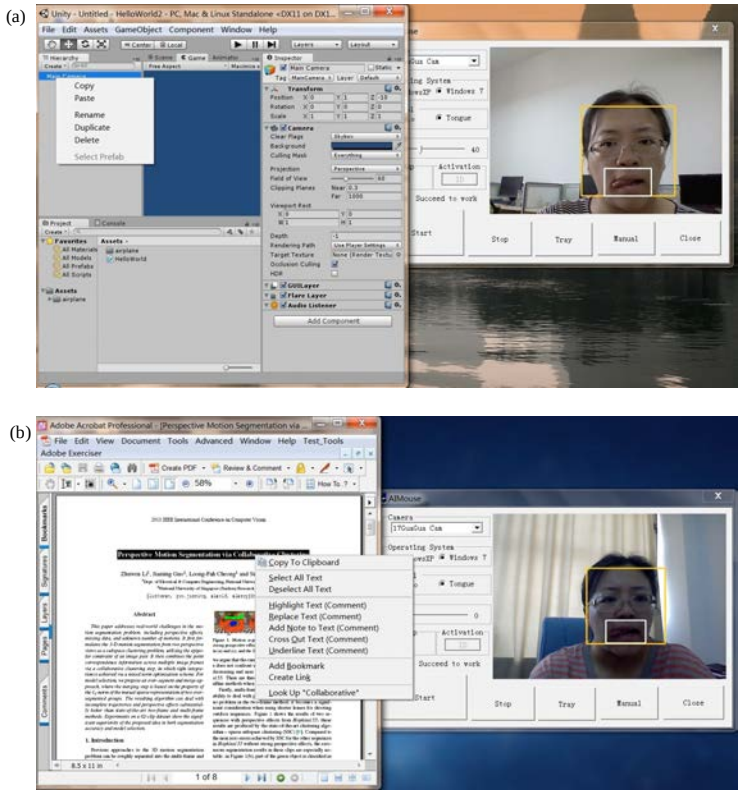

Fig. 7(a-b): Experimental results of sticking out tongue to right in different experimental environments

Table 1: Experimental results of the proposed algorithm

Videos	PR	RE
Test video 1	0.885	0.901
Test video 2	0.830	0.895
Test video 3	0.865	0.921

Figure 8 presents the experiment result of multi-faces in different experimental environments. The system can collect and detect all faces in video frames, but it assumes the face closest to the screen as system operator and only responds to and tracks it. In the operation process, if the face calibrated by the system disappears or removes from the screen, the system message displays: Searching and the function of the mouse suspends, until the system redetects the face image, as shown in Fig. 4b. If the face image pixel detected by the system is less than 100×100, the mouse system will not be started. When the face image is too small or the motions of mouth are not well captured, the system will give up the current frame and proceed to the next one.

The experimental results take detection precision (PR) and recall (RE) rate as evaluation indexes (Su *et al*., 2014) and take occurrences, correct detection frequency and missing detection frequency of the tongue's actions as evaluation elements. Different test results are shown in Table 1. Compared with other method which uses Gabor wavelet (Cao and Wang, 2006) to get the character of the mouth area and uses the SVM classifier for classification, the algorithm detection

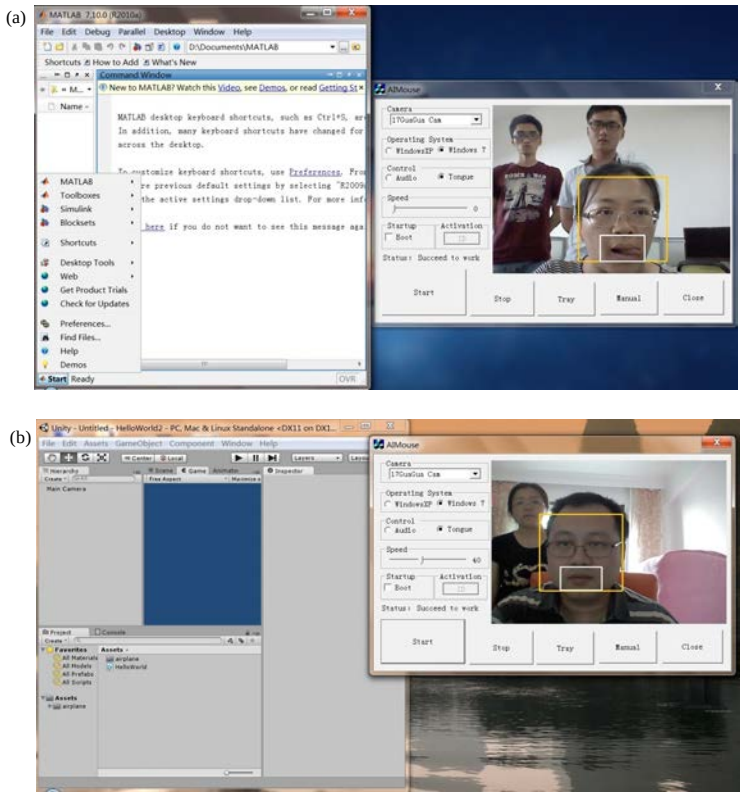

Fig. 8(a-b): Experimental results of multi-faces in different experimental environment

precision reaches 0.657 and the recall rate gets 0.701 in the test of video 1. From the result, it can be figure out that the algorithm put forward in this study is of higher detection precision and recall rate.

Evaluation of real-time performance: According to the experiment, the average detection time of mouse moving in the system is 4 msec while the average detection and activation time of mouse click is 4-8 msec and slight difference exists in the time calculated by different systems. The optical mouse which is currently widely used has the Mouse Rolling Rate (https://wiki.archlinux.org/index.php/Mouse_Polling_Rate) of 125 Hz or 500 Hz and the response delay of 2-8 msec. Therefore, the present system can not only achieve the general mouse usage standard, but also be real-time.

CONCLUSION

A new tongue controlled mouse system based on computer vision was proposed in this study while the face tracking and the pattern recognition technologies were used to implement the operation of video mouse. In addition, we proposed an Advanced LBP operator to extract the tongue feature for identifying the tongue action. Experimental results show that the algorithm completes the detection of tongue action which is of high robustness to meet the needs of real-time applications. However, the feature vector obtained in the algorithm have high dimension, dimensionality reduction and the detection accuracy improvement are the focus of future study.

ACKNOWLEDGMENT

This study is supported by National Natural Science Foundation of China (61300089) and the Fundamental Research Funds for the Central Universities of China (DC12010107).

REFERENCES

Cao, L. and D.F. Wang, 2006. Face recognition based on two-dimensional gabor wavelets. J. Elect. Inform. Technol., 28: 490-494.

Chen, Z.X., Y.M. Cheng and D. Zeng, 2009. Tongue movements and mouth expressions animation based on multi-curve spectrum. J. Syst. Simul., 21: 7518-7521.

Chongqing Academy of Science and Technology, 2013. The implementation method and system of eye controlled mouse. China Patent CN201310130392.9, June 26, 2013.

Hankyong Industry Academic Cooperation Center, 2009. Head mouse. Korea KR20090071335A, August 3, 2009.

Mao, L., 2008. Correction method of face image. CN101163189 A, April 16, 2008.

Ojala, T., M. Pietikainen and D. Harwood, 1996. A comparative study of texture measures with classification based on featured distributions. Pattern Recognit., 29: 51-59.

Poppe, R., 2010. A survey on vision-based human action recognition. Image Vision Comput., 28: 976-990.

Schreiber, D., 2008. Generalizing the Lucas-Kanade algorithm for histogram-based tracking. Pattern Recognit. Lett., 29: 852-861.

Su, Y., A. Li, K. Jiang and G. Jin, 2014. Improved background extractor model for moving objects detecting algorithm. J. Comput. Acided Des. Comput. Graph., 26: 232-240.

Tan, X. and B. Triggs, 2010. Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans. Image Proces., 19: 1635-1650.

Tang, Y., Z. Huang, R. Huang, J. Jiang and X. Lu, 2011. Texture image classification based on multi-feature extraction and SVM classifier. Comput. Applic. Software, 28: 22-25.

Wang, Y., W.H. Li, B. Fu, H.Y. Li and H.Y. Ni, 2011. Face recognition algorithm using RDW-LBP based on horizontal component prior principle. J. Jilin Univ., 41: 750-757.

Wang, L. and Z. Wang, 2012. Application of agpso-based adaboost algorithm in human face detection. Energy Procedia, 13: 1221-1230.

Zhang, Z.W. and H.B. Shen, 2008. Lip detecting algorithm based on chroma distribution diversity. J. Zhejiang Univ., 42: 1355-1359.