International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.10, (2016), pp.205-216
http://dx.doi.org/10.14257/ijsip.2016.9.10.20

Eye-hand Coordination based Human-Computer Interaction

Kang Wei[1] and Ye-peng Guan[1, 2]

[1]*School of Communication and Information Engineering, Shanghai University*
[2]*Key Laboratory of Advanced Displays and System Application, Ministry of Education*
ypguan@shu.edu.cn

Abstract

Human-computer interaction (HCI) has great interactive applications in many fields. A novel eye-hand coordination based HCI approach has been proposed. According to the fact that the eye gazing starts prior to the hand movement and reaches the target in advance, an eye-hand coordination model in a non-contact way is constructed by facial orientation and skeletal joints of hand. Both temporal median filtering and moving average filter strategies are developed to overcome some fluctuation influences during HCI. The diversity of interactive habits among multiple users is considered in an ordinary hardware from a crowded scene without any hypothesis for the scenario contents in advance. Comparative comparisons with state-of-the-arts have highlighted the superior performance of the proposed approach.

Keywords: *HCI; eye-hand coordination; eye gazing; hand pointing*

1. Introduction

HCI is an active field of research in computer vision fields and the information technology for interactive applications [1]. There are many HCI patterns including facial expression, body posture, hand gesture, speech recognition and any other patterns [2]. Face and gesture based HCI have been developed in [3]. The areas of an image using skin colors are labeled which act as candidates for the face and hand. A gesture recognition approach has been proposed to perform HCI in [4]. Some significant features, the quaternions of some skeleton joints are employed and a supervised approach is used to build the gesture models. Elmezain *et. al.,* [5] proposed a HCI approach to executing hand gesture spotting and recognition based on hidden Markov models. Suk *et. al.,* [6] proposed a HCI method based on hand gestures in a continuous video stream using a dynamic Bayesian network model. Yao and Fu [7] developed a hand gesture interaction method based on Kinect sensor. Compared to hand gesture with diversity, pointing gesture can be easily recognized and can be taken as one of more natural human computer interfaces [8]. Park and Lee [9] present a 3D pointing gesture recognition based on a cascade hidden Markov model and a particle filter for interacting with mobile robots. Kehl and Gool [10] proposed a multi-view method to estimate 3D directions of one or both arms. Michael *et. al.,* [11] has set up two orthogonal cameras to detect hand regions, track the finger pointing features, and estimate the pointing directions in 3D space. The pointing direction in [9-11] is determined according to the results of face and hand tracking. The performance of HCI is limited by the unreliable face and hand detection. Another difficult problem is how to recognize some small pointing gestures which usually results in the wrong direction estimation. Pan and Guan [12] proposed an adaptive virtual touch screen to perform HCI to overcome this problem. Although these methods [3-12] have been proven to be useful for HCI, some assumptions such as known interactive user and his corresponding activity must be given in advance which is not practical for realistic HCI scenario. Besides, these methods [3-12] cannot predict efficiently the interactive

ISSN: 2005-4254 IJSIP

intention during HCI, especially there is multiple users with some different interactive gestures at the same time. One important element of automatical recognition of human activity is eye-hand coordination, which is fundamental to daily activities [13] studied in psychology and cognitive science [14-15]. Our eyes and hands move in coordination to execute many everyday tasks. Eye-hand coordination is expressed also in the pointing behavior. Some methods have been developed for HCI based on eye-hand coordination with wearable equipment attached to the human body [16-19]. Both freedom and flexibility for the interactive user have been restrained to a large extent.

A novel eye-hand coordination based HCI approach has been proposed in this paper. According to the fact that the gaze starts prior to the hand movement and reaches the target before hand velocity peak [20], a non-wearable eye-hand coordination model is constructed by facial orientation and skeletal joints of hand. The main contributions are as follows. The first contribution is that an eye-hand coordination model in a no-contact way for HCI has been constructed. The second contribution is that both temporal median filtering and moving average filter strategies are developed to overcome some fluctuation influences during HCI. The third contribution, maybe the most important, is that the diversity of interactive habits among multiple users is considered in an ordinary hardware from a crowded scene without any hypothesis for the scenario contents in advance. Comparative comparisons with state-of-the-arts have indicated the superior performance of the proposed method.

The organization of the rest paper is as follows. In Section 2, eye-hand coordination based HCI is described. Experimental results and analysis are discussed in Section 3 and followed by some conclusions in Section 4.

2. Eye-Hand Coordination Based HCI

Rapid discrete goal-directed movements are characterized by a coordination pattern between the gaze and the hand displacements. When the target is not known in advance, or in the case of sequential aiming, eye and hand movements towards a target synergistically to maximize accuracy, and minimize the temporal and energy costs of the limb in advance of the limb movement [21].

2.1. Gazing Direction Estimation and Gazing Based HCI

People interact naturally with each other by their face to convey visual information. A person's face orientation is an indication of what he or she is the most interested in, or with which his or her interacts. Face orientation can be applied to estimate gaze direction intuitively.

Active shape models (ASM) is one of a popular shape modeling and feature extraction method, which is proposed by Cootes *et. al.*, [22] and developed by other researchers in recent years. ASM [22] is adopted to locate the facial contour feature points for facial orientation instead of AAM [23] due to its less time in fitting.

An OpenNI SDK is used to capture color images, depth images and skeleton data from Kinect sensor at first. The skeleton map consists of 25 distinct and primary skeletal joints of the human body. The head joints captured from skeleton information are used to estimate roughly the face region by detecting the face only in the estimated region of color images. The position and size of the face region is employed to initialize the ASM [22] model parameters. We make use of all the facial contour feature points extracted by ASM [22] in color image acquired from the Kinect to construct a 2D facial geometric model. After calibrating color camera and depth camera of Kinect in [24], the 2D geometric model is mapped to a 3D coordinate space as follows.

Each facial contour feature points in color image are presented by a 3D vector $P_i = [X_i, Y_i, Z_i]$. Label the contour feature points according to the counterclockwise from 0 to (N-1). Based on the N numbered feature points, an initial triangulation divides the contour

feature points into numerous triangles. A chart of modeling facial geometry is shown in Figure 1.

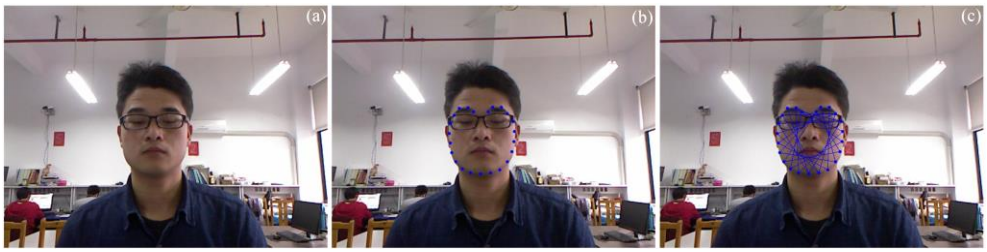

Figure 1. Facial Geometric Modeling. (a) An Original Color Image. (b) Extracted Contour Feature Points. (c)Triangularization of Facial Feature Points

For each triangle, three vertices ***(A, B, C)*** in a triangle are employed to estimate the facial normal direction as following:

$$\overrightarrow{n_f} = \left[X_{nf}, Y_{nf}, Z_{nf}\right] = \overrightarrow{P_A P_B} \otimes \overrightarrow{P_A P_C} \tag{1}$$

where the symbol ⊗ is a cross operator.

The facial interactive point for the origin with the nasal tip representing projection one of gazing is constructed as:

$$G_{ft} = \left[X_G, Y_G, Z_G\right] = P_{nose} + n_0 \overrightarrow{n_f} \tag{2}$$

with respect to

$$n_0 \cdot \overrightarrow{n_f} = n_0 \left[X_{nf}, Y_{nf}, Z_{nf}\right] \tag{3}$$

where P_{nose} is the 3D coordinate of nasal tip point, n_0 is a constant to make Z_G coordinates of the facial interactive point to zero.

Since original depth map has some errors such as holes caused by occlusion, there are some random incorrect depth data [25]. Meanwhile, the facial interactive point will be deviated from each other due to the fluctuation caused by rotation of face. This makes it difficult to determine facial orientation steadily. In order to overcome the problem, a temporal median filter strategy is developed. Several successive frames are used to perform temporal median filter to determine the interactive point as:

$$G_f = \frac{1}{G_n \cdot N_t} \sum_{i=0}^{G_n - 1} \sum_{j=0}^{N_t - 1} G_{ft}(i, j) \tag{4}$$

where $G_{ft}(i, j)$ is an interactive point in the *jth* triangle of *ith* frame, G_n is frame number selected for the temporal median filtering, N_t is total number of triangles used.

Some results for gazing direction estimation and gazing based HCI are shown in Figure 2, from different views.

Figure 2. Gazing Direction Estimation and Gazing Based HCI From Left to Right, Respectively

The yellow line represents the estimated gazing direction; a lamp fixed at the wall is lit by the gazing direction in Figure 2, from left to right, respectively.

2.2. Hand Movement Estimation and Pointing Based HCI

According to the fact that Kinect can be used to track the movements of 25 distinct skeletal joints on the human body, hand tracking and location is realized by continuously processing color image and depth images. After calibration of Kinect [24], the frames of the color camera are aligned with those of the depth images. When user reached out his hand, the 3D coordinates of hand joint can be gotten from the skeletal map. We determine which hand moves to the goal corresponding with the gazing as:

$$H_g = \begin{cases} H_l, & Z_r \geq Z_l \\ H_r, & else \end{cases} \tag{5}$$

where H_r represents the right hand, H_l is the left hand, Z_r and Z_l present the Z coordinates of right and left hand, respectively.

Both hand and elbow skeleton joints are figured out through skeleton-depth conversion based on the skeletal information captured by Kinect. The hand movement direction is determined according the 3D coordinates of hand and elbow joints.

Let elbow and hand joints are represented by 3D vector P= [X_e, Y_e, Z_e] and Q=[X_h, Y_h, Z_h], respectively. The hand movement direction is determined as:

$$\overrightarrow{PQ} = \left(X_e - X_h, Y_e - Y_h, Z_e - Z_h\right) \tag{6}$$

Some results for hand motion estimation and skeletal joint extraction are shown in Figure 3.

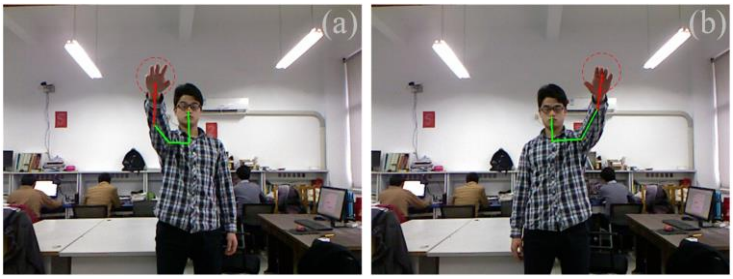

Figure 3. Hand Motion Direction Estimation and Skeletal Joint Extraction

The green lines represent the hand skeleton motions in Figure 3. The red solid line represents hand pointing direction, while the red dashed circle represents hand pointing region.

After we have found the contour of hand, we compute the gradient angle of different pixels in the contour of the hand instead of computing the curvature as:

$$\Phi = \arccos \frac{\vec{a} \cdot \vec{b}}{\left\|\vec{a}\right\| \left\|\vec{b}\right\|} \tag{7}$$

with respect to

$$a = \left(X_{i+1} - X_i, Y_{i+1} - Y_i\right) \tag{8}$$

$$b = \left(X_{i-1} - X_i, Y_{i-1} - Y_i\right) \tag{9}$$

where (X_i, Y_i) is the coordinate of potential fingertip pixel at the *ith* frame.

To overcome the influence of localization for potential fingertip, 3D coordinates of the pixels in the contour from the depth map to determine whether the potential fingertip is true or not:

$$\vec{\mu} = \vec{a}' \times \vec{b}' \tag{10}$$

with respect to

$$a' = \left(X_{i+1} - X_i, Y_{i+1} - Y_i, Z_{i+1} - Z_i\right) \tag{11}$$

$$b' = \left(X_{i-1} - X_i, Y_{i-1} - Y_i, Z_{i-1} - Z_i\right) \tag{12}$$

If the vector μ and Z-axis positive direction are at the same direction, we determine the potential fingertip is a true one.

Both the 3D position of fingertip P_{ftip} and the hand's joint position P_{hand} are used to calculate the hand pointing based interactive point as:

$$H_p = \left[X_{H_p}, Y_{H_p}, Z_{H_p}\right] = P_{ftip} + c_0 \cdot \overrightarrow{n_p} \tag{13}$$

$$P_{ftip} + c_0 \cdot \overrightarrow{n_p} = \left[X_{ftip}, Y_{ftip}, Z_{ftip}\right] + c_0 \cdot \left[X_{hf}, Y_{hf}, Z_{hf}\right] \tag{14}$$

where c_0 is a constant to make Z_{Hp} to zero, and

$$\overrightarrow{n_p} = P_{hand} - P_{fitp} \tag{15}$$

Since there are some jitters for fingertip when the pointing hand moves, a moving average filter strategy is developed to get a stabile interactive point as:

$$p_{af} = \frac{1}{P_n} \sum_{i=0}^{P_n - 1} p_{fi} \tag{16}$$

where $P_{fi}(i, j)$ is fingertip position at the *ith* frame, P_n is frame number selected for the smoothing filter.

Some results for pointing based HCI are shown in Figure 4, from different views.

Figure 4. Pointing Direction Estimation and Pointing Based HCI From Left to Right, Respectively

The red line represents the estimated pointing direction. A lamp fixed at the wall is lit by the hand pointing in Figure 4, from left to right, respectively.

2.3. Eye-Hand Coordination Based HCI

To test whether eye and hand movements towards a target is coordination or not, a cosine similarity measurement strategy is proposed. Let vector $\alpha = (x_g, y_g, z_g)$ and $\beta = (x_g, y_g, z_g)$ are the gazing and hand movement directions, respectively. The cosine value between α and β is calculated:

$$\cos\left\langle \vec{\alpha}, \vec{\beta} \right\rangle = \frac{\vec{\alpha} \cdot \vec{\beta}}{\left\| \vec{\alpha} \right\| \left\| \vec{\beta} \right\|} \tag{17}$$

The degree of eye-hand coordination is computed as:

$$sim\left(\vec{\alpha}, \vec{\beta}\right) = \left| \cos\left\langle \vec{\alpha}, \vec{\beta} \right\rangle \right| \tag{18}$$

For sim with [0, 1], value 1 represents eye and hand movements exactly towards the same target, while 0 indicates that they are independent.

The higher is the sim, the more similar is the goal-directed movements caused by eye and hand. In other words, it can be used to distinguish an interacting user with eye-hand coordination from the HCI scenario with a higher sim. It is crucial how to tell the interactive user from others in the sim. Since there are some differences for both gazing and hand motion directions, the sim is revised as:

$$Sim_{EHC} = \begin{cases} 1 & if \quad sim \geq T \\ 0 & else \end{cases} \tag{19}$$

where T is a threshold value.

It indicates that eye and hand moves to the same target when SimEHC is 1 from (19). Since there some differences between the interactive points with gazing and pointing, Euclidean distance measurement is proposed to determine the interactive point as:

$$S_t = \begin{cases} 1 & if \; D\left(G_f, H_P\right) \leq D_T \\ 0 & else \end{cases} \tag{20}$$

with respect to

$$D\left(G_f, H_P\right) = \sqrt{\left(X_G - X_{Hp}\right)^2 + \left(Y_G - Y_{Hp}\right)^2} \tag{21}$$

where D_T is a threshold value.

Some results for eye-hand coordination based HCI are shown in Figure 5.

Figure 5. Eye-Hand Coordination Estimation and HCI Based on It

The first row images represent the movement trends that eye and hand movements towards a target in Figure 5. The numbers at the top left corner represent the degree of eye-hand coordination. The second row images are the corresponding interactive responses. The lamp is turn on when eye and hand movements towards the same target shown at the right image in the second row in Figure 5.

3. Experimental Results and Analysis

In order to test the performance of the proposed method, some experiments have been done in a platform shown at Figure 6.

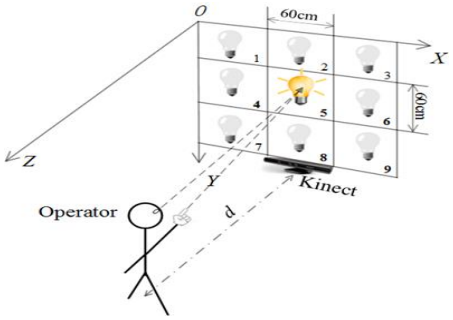

Figure 6. Eye-Hand Coordination Based HCI Platform

The nine lamps at the platform are taken as the interactive targets. Ten volunteers with different interactive habits take part in the tests. The video frame is 640×480 pixels captured by the Kinect. The experiment is performed in an Intel Core I3-2120 CPU with 4GBs RAM in C++.

3.1. Parameter Analysis

Since there are some fluctuations during gazing and hand pointing based HCI, both temporal median filtering and moving average filter strategies have been developed in (4) and (16), respectively. The processing time will be increased for a larger value G_n and P_n, while the fluctuation would not be suppressed for a smaller value G_n and P_n. To get a reasonable trade-off between the efficiency and interactive performance, some variances for the interactive points with different G_n and P_n are given in Figure 7, respectively.

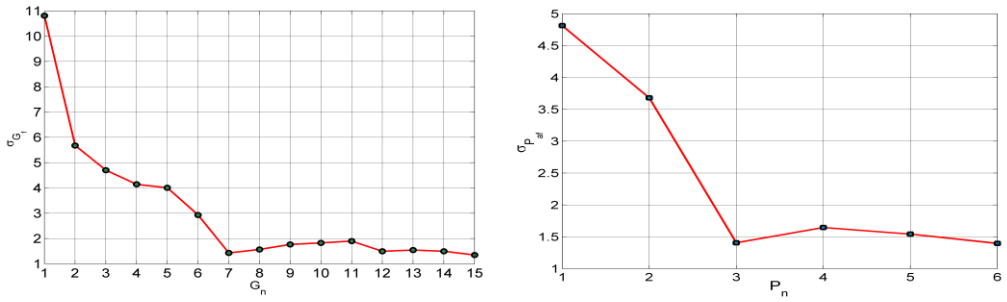

Figure 7. Variance Curves for the Interactive Points With Different Gn and Fn From Left to Right, Respectively

One can note that the variance changes gently after G_n is more than 7 and the variance is minimal when G_n is 7 from the left curve in Figure 7. At the same time, the variance changes gently after P_n is more than 3 and the variance is minimal when P_n is 3 from the right curve in Figure 7. The G_n and P_n are set to 7, 3, respectively, and kept the same in the experiment.

For the threshold values T in (19), and D_T in (20), an average correct recognition rate R_C is computed as:

$$R_C = \frac{N_t - N_w}{N_t} \times 100\% \tag{22}$$

Where N_t is the total true interactive numbers, N_w is the wrong interactive numbers. The average R_Cs with different T and D_T are given in Figure 8.

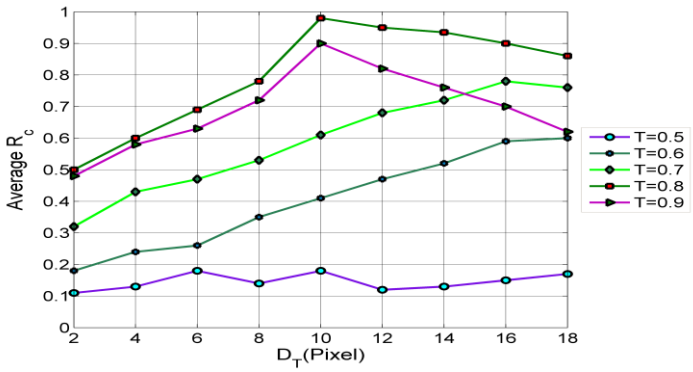

Figure 8. The Average R_C with Different T and D_T

One can find that the interactive performance is the best when T is set to 0.8, and D_T set 10 from Figure 8. The T and D_T are set to 0.8, 10, respectively, and kept the same in the experiment.

3.2. Quantitative Evaluations and Comparisons

To quantitatively evaluate the HCI performance in different interactive ways including gazing, hand pointing and eye-hand coordination, some results in a confusion matrix way are given in Table 1, 2, and 3, respectively. L_1, L_2, L_3, L_4, L_5, L_6, L_7, L_8, and L_9 in the Tables represent the nine interactive lamps on the wall shown at Figure 6, respectively.

Table 1 Gazing Based HCI

	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9
L_1	**95.2%**	2.4%	0	1.8%	0.6%	0	0	0	0
L_2	1.2%	**94.3%**	1.8%	0.6%	1.4%	0.7%	0	0	0
L_3	0	1.6%	**95.5%**	0	0.9%	2.0%	0	0	0
L_4	0.8%	0.7%	0	**95.6%**	1.2%	0	0.8%	0.9%	0
L_5	0.4%	0.8%	0.4%	1.2%	**93.8%**	1.4%	0.2%	1.2%	0.6%
L_6	0	0.6%	1.0%	0	1.6%	**95.4%**	0	0.4%	1.0%
L_7	0	0	0	2.2%	1.6%	0	**93.2%**	3.0%	0
L_8	0	0	0	0.7%	0.9%	0.8%	0.9%	**95.7%**	1.0%
L_9	0	0	0	0	1.0%	1.8%	0	1.8%	**95.4%**

Table 2 Hand Pointing Based HCI

	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9
L_1	**97.3%**	0.9%	0	1.2%	0.6%	0	0	0	0
L_2	0.6%	**96.0%**	0.7%	0.9%	0.8%	1.0%	0	0	0
L_3	0	0.6%	**97.6%**	0	0.6%	1.2%	0	0	0
L_4	0.4%	0.2%	0	**96.8%**	0.6%	0	1.2%	0.8%	0
L_5	0.2%	0	0	0.2%	**97.6%**	0.2%	0.4%	0.8%	0.6%
L_6	0	0.2%	0.4%	0	0.8%	**96.8%**	0	0.6%	1.2%
L_7	0	0	0	0.8%	1.0%	0	**96.6%**	1.6%	0
L_8	0	0	0	0.2%	0.4%	0.8%	1.4%	**95.4%**	1.8%
L_9	0	0	0	0	0.4%	0.8%	0	2.0%	**96.8%**

Table 3 Eye-hand Coordination Based HCI

	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9
L_1	**99.2%**	0.2%	0	0.2%	0.4%	0	0	0	0
L_2	0.2%	**98.8%**	0.2%	0.1%	0.5%	0.2%	0	0	0
L_3	0	0.2%	**99.1%**	0	0.5%	0.2%	0	0	0
L_4	0.1%	0.3%	0	**98.8%**	0.4%	0	0.1%	0.3%	0
L_5	0.3%	0.2%	0.2%	0.2%	**98.0%**	0.2%	0.4%	0.2%	0.3%
L_6	0	0.1%	0.3%	0	0.5%	**98.7%**	0	0.1%	0.3%
L_7	0	0	0	0.3%	0.1%	0	**99.4%**	0.2%	0
L_8	0	0	0	0.3%	0.4%	0.2%	0.3%	**98.6%**	0.2%
L_9	0	0	0	0	0.3%	0.2%	0	0.3%	**99.2%**

One can find that the performance of eye-hand coordination based HCI is the best from the Tables. Moreover, one can note that the developed method has better HCI performances even in single modal including gazing and hand pointing with multiple interactive users at the same time from Table 1, to Table 2, respectively.

In order to evaluate the real-time performance in different interactive ways mentioned above, some results of the time consuming for each frame are given in Table 4. One can find that the time consuming in the eye-hand coordination is better than that in hand pointing from Table 4.

Table 4 Time Consuming Statistics in Different Interactive Modal

Interactive modal	Gazing	Hand pointing	Eye-hand coordination
Time-consuming	79ms	98ms	92ms

One can find that the developed interactive method even in single modal can be applied in real-time HCI from Table 4.

To further evaluate the performance of proposed method, some approaches [3,-12] are selected to test at the same conditions. The experimental results are given in Table 5.

Table 5 Comparisons in Different Methods

Method	[3]	[7]	[12]	**Proposed**
Average R_c	95.8%	96.9 %	97.3 %	**98.9%**
Time-consuming	116ms	87ms	98ms	**92ms**

One can find that the recognition performance of the proposed method is the best among the investigated methods from Table 5. The performance of the time-consuming in the proposed method is approximately equivalent that of one [7], while it is superior to that of ones [3] and [12]. As a whole, the developed approach has the best performance by comparisons.

4. Conclusions

A novel eye-hand coordination based HCI in a non-contact way has been developed. ASM is employed to locate the facial contour feature points. The gazing direction is estimation by facial orientation. Hand tracking and location is realized by continuously processing color image and depth images capture by the Kinect. The hand movement direction is estimated by the skeletal joints of hand. Both temporal median filtering and moving average filter strategies are developed to overcome some fluctuation influences during gazing and hand pointing, respectively. The developed interactive method even in single modal including gazing and hand pointing can be applied in real-time HCI with a better performance. Some methods have been selected to test the HCI performance. The proposed approach has superior interactive performance in an ordinary hardware from a crowded scene without any hypothesis for the scenario contents in advance by comparisons.

In the future, we would do further work in some more robust features to improve the HCI performance in dynamic cluttered scenarios.

Acknowledgments

This work is supported in part by the Natural Science Foundation of China (Grant no. 11176016, 60872117), and Specialized Research Fund for the Doctoral Program of Higher Education (Grant no. 20123108110014).

References

[1] F. Karray, M. Alemzadeh, J. A. Saleh and M. N. Arab, “Human-computer interaction: overview on state of the art”, International Journal on Smart Sensing and Intelligent Systems, vol. 1, no. 1, **(2008)**, pp. 137-159.

[2] R. Harish, S. A. Khan, S. Ali and V. Jain, “Human computer interaction - a brief study”, International Journal of Management, IT and Engineering, vol. 3, no. 7, **(2013)**, pp. 390-401.

[3] Y. J. Tu, C. C. Kao and H. Y. Lin, “Human computer interaction using face and gesture recognition”, Proceedings of IEEE International Conference on Signal and Information Processing, Kaohsiung, China, **(2013)** October 29-November 1.

[4] C. Attolico, G. Cicirelli, C. Guaragnella and T. D'Orazio, "A real time gesture recognition system for human computer interaction", Proceedings of International Workshop on Multimodal Pattern Recognition of Social Signals in Human Computer Interaction, Stockholm, Sweden, **(2015)** August 24.
[5] M. Elmezain, A. A. Hamadi and B. Michaelis, "Hand trajectory-based gesture spotting and recognition using HMM", Proceedings of International Conference on Image Processing, Cairo, French **(2009)** November 7-10.
[6] H. I. Suk, B. K. S. Sin and W. Lee, "Hand gesture recognition based on dynamic Bayesian network framework", Pattern Recognition, vol. 43, no. 9, **(2010)**, pp. 3059-3072.
[7] Y. Yao and Y. Fu, "Contour model based hand-gesture recognition using Kinect sensor", IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, no. 11, **(2014)**, pp. 1935-1944.
[8] C. Colombo, A.D. Bimbo, A. Valli, "Visual capture and understanding of hand pointing actions in a 3-D environment", IEEE Transactions on Systems, Man, and Cybernetics - Part B, vol. 33, no. 4, **(2003)**, pp. 677-687.
[9] C. B. Park and S. W. Lee, "Real-time 3D pointing gesture recognition for mobile robots with cascade HMM and particle filter", Image and Vision Computing, vol. 29, no. 1, **(2011)**, pp. 51-63.
[10] R. Kehl and L. V. Gool, "Real-time pointing gesture recognition for an immersive environment", Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, Zurich, Switzerland, **(2004)** May 17-19.
[11] J. R. Michael, C. Shaun and J. Y. Li, "A multi-gesture interaction system using a 3-D iris disk model for gaze estimation and an active appearance model for 3-D hand pointing", IEEE Transactions on Multimedia, vol. 13, no. 3, **(2011)**, pp. 474-486.
[12] J. Pan and Y. P. Guan, "Human-computer interaction using pointing gesture based on an adaptive virtual touch screen", International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 6 no. 4, **(2013)**, pp. 81-92.
[13] S. Thiemjarus, A. James and G. Z. Yang, "An eye-hand data fusion framework for pervasive sensing of surgical activities", Pattern Recognition, vol. 45, no. 8, **(2013)**, pp. 2855-2867.
[14] B. Rasolzadeh, M. Bjorkman, K. Huebner and D. Kragic, "An active vision system for detecting, fixating and manipulating objects in real world", International Journal of Robotics Research, vol. 29, no. 2-3, **(2013)**, pp. 133-154.
[15] D. Jonikaitis and H. Deubel, "Independent allocation of attention to eye and hand targets in coordinated eye-hand movements", Psychological Science, vol. 22, no. 3, **(2011)**, pp. 339-347.
[16] K. Sakita, K. Ogawara, S. Murakami, K. Kawamura and K. Ikeuchi, "Flexible cooperation between human and robot by interpreting human intention from gaze information", Proceedings of IEEE International Conference on Intelligent Robots and Systems, New Orleans, America, **(2004)** September 28- October 2.
[17] Y. Tamura, J. Ota, T. Arai and M. Sugi, "Estimation of user's intention inherent in the movements of hand and eyes for the deskwork support system", Proceedings of IEEE International Conference on Intelligent Robots and Systems, San Diego, America **(2007)** October 29-November 2.
[18] M. Carrasco and X. Clady, "Exploiting eye-hand coordination to detect grasping movements", Image and Vision Computing, vol. 30, no. 11, **(2012)**, pp. 860-874.
[19] L. Twardon, A. Finke and H. Ritter, "Exploiting eye-hand coordination: a novel approach to remote manipulation", Proceedings of IEEE International Conference on Intelligent Robots and Systems, Kyoto, Japan, **(2013)** November 3-7.
[20] R. Terrier, N. Forestier, F. Berrigan, M. G. Robitaille, M. Lavallière and N. Teasdale, "Effect of terminal accuracy requirements on temporal gaze-hand coordination during fast discrete and reciprocal pointings", Journal of Neuro Engineering and Rehabilitation, vol. 8, no. 1, **(2012)**, pp. 1-12.
[21] G. Binsted, R. Chua, W. Helsen and D. Elliott, "Eye-hand coordination in goal-directed aiming", Human Movement Science, vol. 20, no. 4-5, **(2001)**, pp. 563-585.
[22] T. F. Cootes, C. J. Taylor and D. H. Cooper, "Graham. Active shape models-their training and application", Computer Vision and Image Understanding, vol. 61, no. 1, **(1995)**, pp. 38-59.
[23] T. F. Cootes, G. J. Edwards and C. J. Taylor, "Active appearance models", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23 , no. 6, **(2001)**, pp. 681-685.
[24] D. Herrera, J. Kannala and J. Heikkila, "Joint depth and color camera calibration with distortion correction", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 10 **(2012)**, pp. 2058-2064.
[25] F. Qi, J. Han, P. Wang, G. Shi and F. Li, "Structure guided fusion for depth map inpainting", Pattern Recognition Letters, vol. 34, no. 1, **(2013)**, pp. 70-76.

Authors

Kang Wei, he is a M.D. Candidate of School of Communication and Information Engineering, Shanghai University, P.R. China. He His major research interests include image processing, non-wearable intelligent human-computer interaction.
E-mail: weikangsd@shu.edu.cn.

Ye-peng Guan, he is a Professor in School of Communication and Information Engineering, Shanghai University, P.R. China. He received the B.S. and M.S. degrees in physical geography from the Central South University, Changsha, China, in 1990, 1996, respectively, and the Ph.D. degree in geodetection and information technology from the Central South University, Changsha, China, in 2000. From 2001 to 2002, he did his first postdoctoral research at Southeast University in electronic science and technology. He did his second postdoctoral research at Zhejiang University in communication engineering, and he had been an Assistant Professor with the Department of Information and Electronics Engineering, Zhejiang University from 2003 to 2004. He has published over 120 papers in related international journals and conferences, such as International Journal of Pattern Recognition and Artificial Intelligence, Engineering Applications of Artificial Intelligence, journal of Engineering, Journal of Electronics, IET Image Processing. He also owns more than 20 patents. His research interests include intelligent information perception, digital image processing, pattern recognition, computer vision, and non-wearable intelligent human-computer interaction. E-mail: ypguan@shu.edu.cn.

