HUMAN COMPUTER INTERACTION ALGORITHM BASED ON SCENE SITUATION AWARENESS

Cai Mengmeng^{1,2}, Feng Zhiquan^{1,2} and Luan Min^{1,2}

 ¹School of Information Science and Engineering , University of Jinan, Jinan ,China 250022
 ²Shandong Provincial Key Laboratory of Network-based Intelligent Computing, Jinan, 250022, P.R.
 Corresponding author: Cai Mengmeng, E-mail: 1414663370@qq.com

Corresponding author: Feng Zhiquan, E-mail: ise_fengzq@ujn.edu.cn

ABSTRACT

Implicit interaction based on context information is widely used and studied in the virtual scene. In context based human computer interaction, the meaning of action A is well defined. For instance, the right wave is defined turning paper or PPT in context B, And it mean volume up in context C. However, Select object in a virtual scene with multiple objects, context information is not fit. In view of this situation, this paper proposes using the least squares fitting curve beam to predict the user's trajectory, so as to determine what object the user's wants to operate .And fitting the starting position of the straight line according to the change of the discrete table. And using the bounding box size control the Z variable to move in an appropriate location. Experimental results show that the proposed in this paper based on bounding box size to control the Z variables get a good effect; by fitting the trajectory of a human hand, to predict the object that the subjects would like to operate. The correct rate is 88.6%.

KEYWORDS

Least-squares method; gesture recognition; human-computer interaction; visualization; implicit interaction; Context information

1. INTRODUCTION

With the continuous development of computer science and technology, intelligent humancomputer interaction has gradually become the dominant trend in the development of computing model. And this trend becomes more obviously after Weiser Mark [1] putting forward the concept of "Ubicomp" in 1990s. In order to lighten the load of people's operation and memory, during the interaction, the traditional way of interaction need to be expanded. And integrate the implicit human-computer interaction into the explicit human-computer interaction.

At present, implicit human-computer interaction has become an important research frontier in the field of interaction. The universities and research institutes of the United States, Germany, China,

David C. Wyld et al. (Eds) : CSITA, ISPR, ARIN, DMAP, CCSIT, AISC, SIPP, PDCTA, SOEN - 2017 pp. 51–66, 2017. © CS & IT-CSCP 2017 DOI : 10.5121/csit.2017.70106

Austria and so on , has been carried out in-depth study to IHCI theory and application gradually. Schmidt in the University of Karlsruhe in Germany conducted an earlier study of the theory of implicit interaction [2].He believes that the two elements of implicit interaction are perception and reasoning, and he also put forward that contextual information is very important for interaction. Hamid Mcheick [3] presents a context aware model with ability to interact. This model adapt to dynamically environment and can interact with the user flexibility. The implicit interaction based on context is also applied in other aspects. Young-Min Jang [4] proposed a novel approach for a human's implicit intention recognition based on the eyeball movement pattern and pupil size variation. Bojan Blažica [5]introduces a new more personal perspective on photowork that aims at understanding the user and his/her subjective relationship to the photos. It does so by means of implicit human-computer interaction, this is, by observing the user's interaction with the photos.

In China, Tao Linmi [6] of Tsinghua University developed an adaptive vision system to detect and understand user behaviour, and to carry out implicit interaction. At the same time, Tian Feng in software research institute of Chinese Academy of Sciences also studied the characteristics of implicit interaction from the perspective of post WIMP user interface [7]. Wang Wei proposes that more use of user context information in the process of implicit human-computer interaction [8], Including user behaviour, emotional state (for example: The emotional design method of Irina CRISTESCU[9]), and physiological state. But there is also some use of Environmental context information, such as location-based services, etc. And it pointed out that the implicit human-computer interaction technology is one of the development directions in the future. Gao Jun pointed out in the article[10] Semantic Analysis is the importance and difficulty of high-level interpretation in image understanding, in which there are two key issues of text image semantic gap and text description polysemy. Yue Weining [11] proposed a context aware and scheduling strategy for intelligent interactive systems, which improves the system's intelligence. And Feng Zhiquan [12] uses the context information in the gesture tracking, and has achieved a good interaction effect.

2. RELATED WORK

52

2.1. Image segmentation

Before image segmentation, the image should be filtered to remove the noise. At present, the common methods of image segmentation [13] can be divided into: Threshold segmentation method [14], edge detection method [15], region segmentation method and the method of combining the theorem of the segmentation method. Besides, Qu Jingjing[16] proposed the segmentation method of continuous frame difference and background subtraction. This article uses the skin colour model [14] (YCbCr) to separate the human hand and the background, and the image banalization. Segmentation results are shown in Figure 1:



Figure 1. Original image and segmented image

2.2. Feature Extraction

The method of feature extraction is varied. Tao Sangbiao [17] proposed a static gesture contour feature extraction algorithm based on contour and skin colour. It extracts the gesture contour though skin's colour, and then extracts contour information. ZHU Jiyu [18] proposed a novel gesture segmentation algorithm can be divided global and local features. A fuzzy set is used to describe the background, colour and motion of the spatial and temporal information in the video stream. Ren Haibing [19] used a variety of information such as colour, motion and edge to extract features that can reflect the structure characteristics of the human hand, And he divides the characteristic lines into small curve segments. and track the movement of these curve segments. Feng Zhiquan [20] proposed the gesture features extraction. This method is not only simple but also has certain rotation and scaling invariance. In this paper, we use the method of document [20] to extract the feature points of the hand gesture. The specific methods as follows:

First, get the segmentation of the coordinates of the hand gesture, and the point of the greatest distance from the coordinates. Second, I using the centroid point as the centre point and the centroid of the farthest point distance concentric circle radius, divided into 7 layers as show in figure 2. Third, these 7 layers are divided into 3 categories: Fingertip layer, Finger heel layer, Joint point layer. In the end, get the fingertip and the number of layers and the number of connectivity.



Figure 2. Feature Extraction

2.3. Gesture Recognition

Hand gesture recognition methods include: template matching method, statistical recognition method, fuzzy recognition method and artificial neural network classification, and shape matching method. Commonly used in the shape matching method [21] has the invariant moment method, the geometric parameter method, the characteristic model representation, the boundary direction histogram method, the wavelet importance coefficient method, as well as our country scholar studies the wavelet contour representation and so on. The method of gesture recognition based on Hausdorff distance [22] template matching algorithm is used in this paper. It is to obtain the characteristics of the library files and calculate the Hausdorff distance, the smaller the distance, the better the matching of the feature points. Specific algorithms are as follows: Assuming that A, B for the two sets has N and M elements respectively, then The distance Hausdorff(A, B) between A and B is H(A, B)

$$H(A, B) = \max\{h(A, B), h(B, A)\}$$
(formula 1)

```
Int Temp = Cnt =0;
```

```
For i=0: N
```

```
For j=0: M

Temp = \min \|a_i - b_j\|;
Cnt = max{Temp, Cnt};

h(A, B) = Cnt;
```

In the same way, you can calculate thus obtained H(A,B).

3. SCENE MODELLING

3.1. Brief introduction of image display

The principle of image display using OpenGL [23] in the virtual environment is exemplified here in Figure 3.



Figure 3. The principle of OpenGL image display

For different Z plane(Z=C, C is a constant), moving the same distance in one condition while the output of distance is not the same (The closer to the point of view, the greater the moving distance on the screen is). Therefore, objects at different coordinates in the virtual scene needs different functions to move them. Moreover, two-dimensional image obtained by the common camera is not good at controlling the movement of three - dimensional hand in three dimensional spaces. So many researchers have used animation as a method to avoiding this problem. Using the principle that the bounding box size is proportional to the image display is the key to control changes in the Z-axis coordinate.



Figure 4. Camera image acquisition principle

The captured image (the size of determined acquisition: 400x300) is mapped to a window to display. So the length of a in plane S1 is $\frac{W_2}{W_1}$ times in plane S2 at the time of display.

3.2. Determine the Relationship of Mapping

Collect and record the size of the bounding box and get its average size when each of the experimenter is operating in the 3D scene. And Mapping shown by MATLAB is shown in Figure 5.



Figure 5. The Height and Width of Gesture Bounding Box

According to the probability formula in Statistics:

$$L = \frac{1}{n} \sum_{i=0}^{n-1} I_i$$
 (formula 2)

$$W = \frac{1}{n} \sum_{i=0}^{n-1} W_i \qquad (formula 3)$$

Calculate the initial value of L and W, $L_0 = 110$, $W_0 = 80$.

Calculate the corresponding relationship between real and virtual human hand (when the bounding box keeps the same size).

$$\frac{\Delta L}{\Delta L} = \frac{400 - 2 * L_0}{L_0} = 0.18125$$
 (for *mula* 4)
$$\frac{\Delta W}{\Delta W} = \frac{300 - 2 * W_0}{W_0} = 0.1833$$
 (for mula 5)

In the virtual scene, $L_0 = 1600$, $W_0' = 1200$. Real hand moves in the horizontal direction a unit, virtual hands should move 5.51 units; and moving a unit in the vertical direction means virtual hands should move 5.45 units. For other positions virtual hands should move $5.51 \frac{L}{L_0}$

units; virtual hands should move $5.45 \frac{W}{W_0}$ units.

For Z coordinates, Position of each object in virtual scene is in [20 30]. The variation of bounding box's length is 80Pixel to 130Pixel. So the congruent relationship of bounding box's length and Z coordinates is:

$$\frac{Z - 20}{L - 80} = \frac{30 - 20}{130 - 80} = 0.2$$
 (formula 6)

That is:

$$z = 20 + (1 - 80) * 0.2$$

4. INTERACTION ALGORITHM BASED ON SCENE SITUATION AWARENESS

4.1. Based on Least Square Method [24] to fit the Motion Trajectory (Broken Line Segment) Algorithm

In order to better fit the motion trajectory of the hand gesture, in this paper, the least square method is used to fit the nonlinear equation.as shown in formula 7:

56

$$y_i = a * x_i + b * \sin(x_i) + c$$
 (i = 1,2,3,..., n) (formula 7)

Formula (x_i, y_i) is the observation coordinate, a is first-order coefficients, b is sine coefficients, and c is constant. a, b and c is the parameter to be solved, assume a_0 , b_0 , c_0 for their approximate value. Order:

$$a = a_0 + \delta a$$
, $b = b_0 + \delta b$, $c = c_0 + \delta c$

Taking y as the dependent variable and X as the independent variable, the error equation is:

$$\mathbf{v}_{yi} = \begin{bmatrix} x_i & \sin(x_i) & 1 \end{bmatrix} \begin{bmatrix} \delta \mathbf{a} \\ \delta b \\ \delta c \end{bmatrix} + a_0 * x_i + b_0 * \sin(x_i) + c_0 - y_i \text{ (formula 8)}$$

Error equation matrix can be expressed as:

$$A\delta X = I + V \qquad (formula 9)$$

Among them:

$$A = \begin{bmatrix} x_1 & \sin(x_1) & 1 \\ x_2 & \sin(x_2) & 1 \\ \mathbf{M} & \mathbf{M} & \mathbf{M} \\ x_2 & \sin(x_n) & 1 \end{bmatrix}, \quad \delta X = \begin{bmatrix} \delta a \\ \delta b \end{bmatrix}$$
$$I = \begin{bmatrix} a_0 x_1 + b_0 \sin(x_1) + c_0 - y_1 \\ a_0 x_2 + b_0 \sin(x_2) + c_0 - y_2 \\ \mathbf{M} \\ a_0 x_n + b_0 \sin(x_n) + c_0 - y_n \end{bmatrix}, \quad V = \begin{bmatrix} V_{y1} \\ V_{y2} \\ \mathbf{M} \\ V_{yn} \end{bmatrix}$$

According to the least square rule (formula 10), fitting a straight line.

$$\sum_{i=1}^{n} \left\| ax_{i} - b\sin(x_{i}) + c_{0} - y_{i} \right\|^{2} = \min \qquad (formula \ 10)$$

And Dependent variable residual is:

$$V = A\delta X - 1$$
 (formula 11)

Because the cycle of $sin(x_i)$ is $2 * \pi$, $bsin(x_i)$ is periodic oscillation among in [0 400], so equation of a curve is

$$y_i = ax_i + b\sin(0.01x_i) + c$$
 $(i = 1, 2, ..., n)$

In the end, According to the coefficient to confirm the good and bad fit.

4.2. Scene Situation Awareness and Interaction Algorithm

Calculate the size of the bounding box, and determine the corresponding relationship. According to the moving direction and distance of the 3D human hand of the two frame image, the movement of the centroid of the human hand is determined. The feature data of the multi frame images is used to synthesize nonlinear curve to predict the direction of human hand movement. And then determine the object at the direction and get the distance to human hand. Therefore, perform the corresponding operation; the specific algorithm is as follows:

First step: Capture a RGB image using a common camera .The height of the image is 400, and the width is 300. Then carry out image segmentation, and image banalization.

Second step: According to the formula (12) of the centroid of mass coordinates [25]:

$$r_c = \frac{\sum_i m_i r_i}{\sum_i m_i}$$
(formula 12)

Figure out the centroid of coordinates after banalization; According to the formula-13 figure out bounding box size.

$$X_{I} = \min_{\substack{f(x_{i}, y_{i}) \neq 0}} \{x_{i}\}$$
 (formula 13 - 1)
$$X_{r} = \max_{i} \{x_{i}\}$$
 (formula 13 - 2)

$$Y_{I} = \min_{\substack{f(x_{i}, y_{i}) \neq 0 \\ f(x_{i}, y_{i}) \neq 0}} \{y_{i}\}$$
(formula 13 - 3)

$$Y_r = \max_{f(x_i, y_i) \neq 0} \{y_i\}$$
 (formula 13 - 4)

 X_i is the left edge of the bounding box, X_r is right edge; Y_i is the upper boundary of the bounding box, Y_r is lower boundary. $f(x_i, y_i) \neq 0$ means that the pixels of the (x_i, y_i) coordinates are skin colour.

Third step: Calculate the vector (the size and direction) between two different centroid of coordinate and determine the direction and distance of the human hand movement in the 3D virtual scene according to the size and coordinates of the bounding box.

$$(\delta x, \delta y) = (x_{i+1}, y_{i+1}) - (x_i, y_i) = (x_{i+1} - x_i, y_{i+1} - y_i)$$

Fourth step: Using the glTranslatef (Dx, Dy, Dz) belonging to OpenGL to change the movement of the three-dimensional human hand in the virtual environment. If the moving amount of one direction (assumed to be X axis direction) is much greater than the other direction (Y axis) so you can only consider the direction where the moving amount is larger.

Fifth step: determine whether frames is greater than a threshold(set to 10). If less than, then return to the first step; Else, use the least square method to simulate curve.

Sixth step: Judge whether the fitting is good. If good, go to step seven; if not, adjust dynamically the number of the current frame according to the change of the discrete table, return to the fourth step.

Seventh step: Determine the number of objects that are in the prediction direction; if there is only one: move the object to the human hand. If not, adjust dynamically the number of the current frame according to the change of the discrete table, return to the fourth step.

At last, carry out the corresponding operation on the object by identifying a series of actions, for example: rotation, scaling, translation, and so on.

Algorithm flow chart is shown in figure 6.



Figure 6. Algorithm flow chart

5. EXPERIMENTAL RESULT

The experiment is divided into two parts. A part is to be familiar with the experimental environment, Operation method and procedure, to determine the mapping relationship. Experimental interface is shown in Figure 7.



Figure 7. Virtual scene

On the right is a virtual 3D scene. Scene consists of virtual hands, and small balls, cylinder, cone and other three-dimensional objects. Each object is fixed, and is not in a z plane. On the left, there have two pictures, one picture is the original, and the other is the split hand. Real hand and virtual hand there is a certain relationship.

I find 65 students to do the experiment in the laboratory environment, under the constant light environment, the completion of the virtual scene to grab objects A, B, C, D, and other simple operation of the experiment. I recorded the size of their gestures when they were in the experiment, calculate the average and mapped with MATLAB, as shown in the figure 5. Determine the corresponding relationship and discrete table data, the content of the discrete table is related to the size of the bounding box and the speed of motion.

Another part is to select the object in the virtual scene as show in figure 7, and then do grab translation and other movements. First, I find 66 students again, divided into equal groups: A team, B team. Secondly, it is clear to tell the A team members of the experimental content: the object of the movement, the speed of movement, etc. Wait until the A team members are all familiar with the experimental environment and operating procedures, to do the experimental data into MATLAB to fit the curve. According to the characteristics of the trajectory of human hand, I fit a curve. As shown in the figure 8.



Figure 8. matlab fitting curve

The picture on the left is the fit figure of all members of the team A grab objects one. In the picture, the blue points are the centres' of Actual trajectory of hand and the red curve is the curve after fitting. The picture on the right is one of them in the left. The curve fitting coefficient is shown in figure 9. By analysing the motion trajectory of the A team, we can see that the trajectory of the hand is similar, and the movement of the hand tends to be circular. According to the trend of curve we can general position of object. We can be seen through the figure 9, the average fitting coefficient is higher than 0.95. This indicates that the selected curve is appropriate. In addition, the blue point is relatively dense in the upper right. That means it will cast lots of time to collision detection.



Without specifying which object to select, Let the B team complete the experiment in the case of prediction and no prediction. Tell everyone select the same object twice. Record the number of

frames for each person to complete the experiment. In the experimental process with predictive function, according to the predicted result, it is judged which object in the virtual scene is to be selected. It will change the object colour and save the current number of frames. Wait the end of the experiment, record forecast result for the wrong. Repeat the experiment 5 times per person, seek its average. The final results are shown in Table 1.

Prediction Correct	1	2	3	4	5
The number of people	0	0	2	33	30

The 6 picture in Figure 10 is the screen in the experimental process with predictive function.



Figure 10. Grab the red ball

In the experimental process without predictive function, Use the method of Team A member to select the object to complete the experiment. Save the feature data and the number of frames used to complete the experiment.

62





Figure 11. Time comparison chart

In Figure 11 we can see B team members to complete the experiment in the experimental process without predictive function; the average number of frames required is greater than 25. In the experimental process with predictive function, the average number of frames required is about 17. And apart from special case, the number of frames required to complete the experiment are within 20. In addition to through the table 1 we can be seen Most people can predict success 5 times, and a little of people predicted success 3 times, nobody can predict success less than 3 times. So, we can get a conclusion: in a specific virtual environment, the use of curve fitting method can be very good to predict the subjects want to operate the object.

6. CONCLUSIONS

According to the movement characteristics of the human hand and people's behaviour habits in the real scene, this article uses the least square method fitting a curve to predict direction of hand movement. This method has achieved very good results. And it can greatly reduce the time of selection. Secondly, the size of the bounding box is used to control the change of the Z axis variable in the appropriate range, and to realize the real manual control of the virtual human hand movement in the three-dimensional space. It conforms to the people in the three-dimensional environment in the operation of the habit (hand before and after the change, the virtual hand before and after the move). It also achieved good interaction effects. Finally, it is opposed to achieve human-computer interaction. And it has a certain effect. But for implementation a more intelligent human-computer interaction, there are a lot of problems to be solved. For example: the speed of the active object near the person, There is occlusion problem, as well as the computer automatically judge whether people have the purpose of the operation, etc.

ACKNOWLEDGEMENTS

This work is supported by National Natural Science Foundation of China under Grant No. 61472163, partially supported by the National Key Research \& Development Plan of China under Grant No.2016YFB1001403, Science and technology project of Shandong Province under Grant No. 2015GGX101025

REFERENCES

- Lee, S.hyun. & Kim Mi Na, (2008) "This is my paper", ABC Transactions on ECE, Vol. 10, No. 5, pp120-122.
- [2] Weiser Mark . The computers for the twenty-first century [J].Scientific American, 1991, 265(3) : 94-104
- [3] Schmidt A. Implicit human computer interaction through context [J] .Personal Technologies, 200, 4 (2/3):191-199.
- [4] Hamid Mcheick. Modeling Context Aware Features for Pervasive Computing [J]. Procedia Computer Science, 2014, 37.
- [5] Young-Min Jang, Rammohan Mallipeddi , Sangil Lee, Ho-Wan Kwak, Minho Lee. Human intention recognition based on eyeball movement pattern and pupil size variation [J]. Neurocomputing, 2013.
- [6] Bojan Blažica, Daniel Vladušič, Dunja Mladenić. A personal perspective on photowork: implicit human-computer interaction for photo collection management [J]. Personal and Ubiquitous Computing, 2013, 178
- [7] Wang Guojian, Tao Linmi. Distributed Vision System for Implicit Human Computer Interaction [J]. Journal of Image and Graphics, 2010,08:1133-1138
- [8] TIAN Feng, DENG Changzhi, ZHOU Mingjun, et al. Research on the implicit interaction characteristic of Post-WIMP user interface. Journal of Frontiers of Computer Science and Technology, 2007, 1(2): 160-169.
- [9] WANG Wei, HUANG Xiaodan, ZHAO Jijun, et al. Implicit Human_Computer Interaction [J]. Information and Control, 2014, 01:101-109.
- [10] Irina CRISTESCU. Emotions in human-computer interaction: the role of nonverbal behaviour in interactive systems [J]. Informatica Economica Journal, 2008, XII2:.
- [11] GAO Jun, XIE Zhao, ZHANG Jun, et al. Image Semantic Analysis and Understanding A Review [J]. Pattern Recognition and Artificial Intelligence, 2010, 02:191-202.
- [12] Yue Weining, WangYue, Wang Guoping et al. Architecture of Intelligent Interaction Systems Based on Context Awareness [J]. Journal of computer-Aided Design and Computer Graphics, 2005, 01:74-79.
- [13] Feng ZQ, Yang B, Zheng YW, Xu T, Tang HK. Hand tracking based on behavioural analysis for users. Ruan Jian Xue Bao/Journal of Software, 2013,24(9):2101-2116 (in Chinese). http://www.jos.org.cn/1000-9825/4368.htm
- [14] S. M. Lock, D. P. M. Wills. VoxColliDe: Voxel collision detection for virtual environments[J]. Virtual Reality, 2000, 51: .
- [15] Haokui-tang. Study of Skin Segmentation Based on Double-Models [D]. Shandong University, 2009.
- [16] Lu Kai, Li Xiaojian, Zhou Jinxing. Hand Signal Recognition Based on Skin Colour and Edge Outline Examination [J]. Journal of North China University of Technology, 2006, 03:12-15.

- [17] QU Jing-jing, XIN Yun-hong. Combined Continuous Frame Difference with Background Difference Method for Moving Object Detection [J]. Acta Photonica Sinica, 2014, 07: 219-226.
- [18] Tao Sangbiao, Jiao Guotai. Study on Extraction Algorithm for Static Hand Gesture Contour Feature [J]. Shanxi Electronic Technology, 2015, 02:90-91.
- [19] ZHU Ji-Yu, WANG Xi-Ying, WANG Wei-Xin, et al. Hand Gesture Recognition Based on Structure Analysis [J]. Chinese Journal of Computers, 2006, 12:2130-2137.
- [20] REN Hai-bing, XU Guang-you, LIN Xue-yin. Hand Gesture Recognition Based on Characteristic Curves [J]. Journal of Software, 2002, 05:987-993.
- [21] FENG Zhi-quan, YANG Bo, ZHENG Yan-wei, et al. Gesture features detection based on feature points distribution analysis [J]. Computer Integrated Manufacturing Systems, 2011, 11:2333-2338+2340-2342.
- [22] Li Junshan, Li Xuhui, Digital Image Processing [D]. Bei Jing, tsinghua university press, 2007 : 264
- [23] ZHANG Liang-guo, WU Jiang-qin, GAO Wen, et al. Hand Gesture Recognition Based on Hausdorff Distance [J]. Journal of Image and Graphics, 2002, 11:43-49.
- [24] kandyer. OpenGL Transform (EB/OL). http://blog.csdn.net/kandyer/article/details/12449973, 2016 -01-18.
- [25] School of Geodesy and Geomatics of Wuhan University. Error theory and measurement adjustment [M].Wuhan: wuhan university press, 2003.
- [26] Zhang Mengzhong. The formula of centroid is derived by mathematical induction. [J].Journal of Jiujiang Teacher's College, 2002, 05: 46-4

AUTHORS

Cai Mengmeng

Master's degree of University of Jinan The main research direction is Human-computer interaction and virtual reality. E-mail: 1414663370@qq.com



Zhiquan Feng

Feng Zhiquan is a professor of School of Information Science and Engineering, Jinan University. He got the Master degree from north-western polytechnica university, china in 1995, and Ph.D degree from Computer Science and Engineering Department, shandong university in 2006. He has published more than 50 papers on international journals, national journals, and conferences in recent years. His research interests include: human hand tracking/recognition/interaction, virtual reality, human-computer interaction and image processing.



Luan Min

Master's degree of University of Jinan. The main research direction is Human-computer interaction and virtual reality. E-mail: 1562920346@qq.com



66