Inference of User's Intention for Human - Robot Cooperation Based on Machine Vision

Bei-Bei YANG, Ge LIU, Xiao-Fan HE, Di ZHAO*
The School of Mechanical Engineering of Hubei University of Technology, 430068 Wuhan, China
*646605009@qq.com

Keywords: Natural human-computer interaction, Intention prediction, Weight distribution, Dominant factor, Machine vision.

Abstract. An intention prediction algorithm for natural human-computer interaction based on machine vision is proposed in this paper. Firstly, the motion data of human skeletal feature point is acquired. Then, the motion data and the real-time interactive image are coupled through data processing. Meanwhile, an intention recognition model for natural human-computer interaction is built based on target feature extraction. The dominant feature weight of the operator intention is distributed by hierarchical method. A parallel scheme is adopted in this algorithm for operator intention recognition. Through experiment and data analysis, the algorithm is proved to be reliable and instrumental for improving the efficiency of natural human-computer interaction.

Introduction

The home service robots which are commonly used for the company of the elderly and the disabled are required to complete the tasks of transferring items, opening doors, identifying and predicting human's intentions, as well as contributing to the cooperation with human and helping the elderly and disabled people with their daily work. A series of methods have been put forward through massive research work by scholars to predict the user's operation intention and facilitate with operation on service robots. For example, De Carli D. and Zhu C. proposed methods to infer the user's intention by the Hidden Markov Model(HMM)[1,2] .J. Elfring adopted the Growing Hidden Markov Model to forecast human's intended position[3].Karim A. Tahboub established Intention- Behavior-State model and introduced dynamic Bayesian network for intention conjecture[4]. The user's interaction intention with service robots in daily application can be described by a variety of ways, such as the body posture, gestures, voice commands, facial expressions or measurement of physiological information like heart rate, skin reactions, etc. Due to the one-way interaction with human and a lack of communication and learning ability, the robots with traditional interaction mode cannot cooperate well with users. Human intention evaluation acting as a natural interaction method, enables robots to understand human intentions and keep learning. Lots of research achievements have been achieved in human's coordination with the computer or robot through user's intention evaluation algorithm. Z. Wang and K Mülling proposed a latent variable model and inferred the intention through the observation of human motion[5]. Graham C. and Juan C. proposed approximate E-M algorithm aimed at the parameters and state assessment of nonlinear stochastic model[6].C.Morato used the extended Kalman filter and combined with multiple Kinect sensor to evaluate human bone structure[7].These algorithms did not provide inference about behavioral intention. Wang Lei put forward a secondary assessment method, which was based on the behavior sequence estimation and behavior rules fitting, to facilitate the multi-robots system with human intention recognition[8].P. A. Lasota put forward a robot speed control algorithm through the research on robot joint angle values and the accurate position of users acquired by the optical camera[9].Harish chaandar Ravichandar and Ashwin Dani proposed the method of using multi-models to infer the human intention[10].

In this paper, the target position which the user's hand would reach was taken as his intention. For example, when the user moves his arms, he wants the robot to predict his target position and adjust the corresponding action. Aimed at predicting the target position, an intention inference algorithm is

ultimately proposed based on a research about the collected human skeleton data and the model built for the human arm movement. The principle of this algorithm is shown in Figure 1.

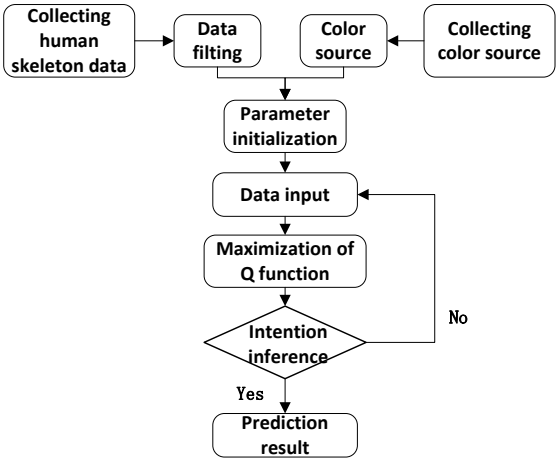


Figure 1**.** Intention inference algorithm

System Modeling

Intention Model: The goal location that the user will reach is taken as the intention model by collecting the dynamics of human arm motion. In this paper, The $d \in D$ was denoted as destination.

$$\mathrm{D} = \{\mathrm{d}_1, d_2, \ldots d_i\} \tag{1}$$

d_i represented the i^{th} destination on the desktop, and $d_i \in R^4$ represented the shape and 3D coordinates location of the destination.

The true intention d can only represent one of the finite destination on a table. The state a_t represented gestures, speed and distance between hand and goal location to describe the position of the arm at a given time t . $\beta_t \in R^3$ represented the measured spatial position of the human arm joint.

Human Arm Action Model: Based on analysis of human motion. It can be obtained that the arm movement was driven by shoulder joint, elbow joint, wrist joints and middle finger joint successively,which caused middle finger joint movement ultimately.

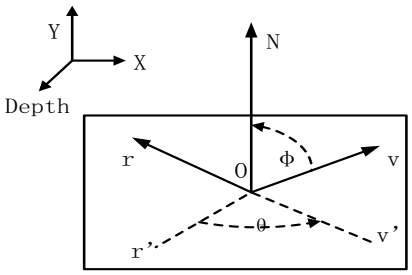


Figure 2. Partial spherical coordinates

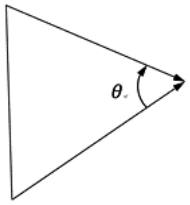


Figure 3. Elbow joint in plane

The shoulder joint has three degrees of freedom. The elbow joint has one degree of freedom. The wrist joint has one degree of freedom. The middle finger joint has one degree of freedom. Thus there are six degrees of freedom in total. The shoulder joint can be represented by a local spherical coordinate system with three variables, namely, azimuth, elevation and radial distance, nevertheless, the radial distance was not used. According to the definition of spherical coordinates, the azimuth angle θ was defined as the angle of counterclockwise rotation ($or' - ov'$). And the elevation angle ϕ is defined as the angle from ov to the vertical axis on ($ov - on$) in Figure.2. The angle of the shoulder joint was defined as:

$$Sh = \{\theta_0, \phi\} \tag{2}$$

With regard to the elbow joint, wrist joint and middle finger with one degree of freedom, the plane of the joint angle should be specified especially. The articulation angle of the elbow joint was defined in Figure. 3. A simplified diagram of the human body's arm motion was shown in Figure. 4.

$$\begin{aligned} EL &= \{\theta_1\} \\ WR &= \{\theta_2\} \\ HA &= \{\theta_3\} \end{aligned}$$

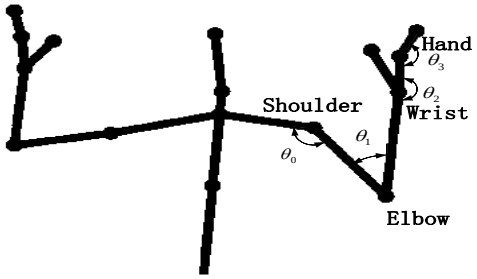


Figure 4. Human arm motion model

According to the measurement of the space coordinates of the human body joint, the angle of each joint was calculated. The range of motion of the arm joint was shown in Table 1.

Table 1. Arm joint Angle range

angle	θ_0	θ_1	θ_2	θ_3
range	-90° $+90^\circ$	0 $+145^\circ$	-80° $+90^\circ$	-30° $+90^\circ$

Intention Inference

The target which the user would reach for was predicted in this paper. There are many ways to express user's intention, for example, the gestures of grabbing different targets, the grasping vector direction of hand, and the distance of the hand from the target object. The intention inference of user from the following four aspects was mainly studied in this paper. Assuming there are n objects on the table, the grasping direction of hand is denoted as

$$\mathrm{V} = \{\mathrm{V}_1, V_2, \ldots V_n\} \tag{3}$$

Due to the different shape of targets, grasping gestures are significantly different. Therefore the four most common hand gestures were defined in the text, which was shown in Table 2.

Table 2. Four types of gestures

Gesture	o	p	l	c
				
Sign	0	1	2	3

The distance between the hand and each possible target object during the movement of the arm is defined as:

$$\mathrm{L} = \{\mathrm{I}_1, \mathrm{I}_2, \ldots \mathrm{I}_n\} \tag{4}$$

The weight of V, G and L are assigned by analytic hierarchy process(AHP) in the intention inference algorithm. According to statistics, when L was far away, V and G can express the intention of the user; when the L was closer, V and L played a dominate role in the user's intention. Therefore, the weight of three kinds of dominant factors are matched based on the variations of L.

S is Relatively Large: According to the rule of pairwise comparison in Table 3, the judgment matrix was shown in Table 4.

Table 3. The meaning of the scale

Proportional scale	Meaning
1	Both elements have the same importance
3	The former two elements than the latter is slightly important
5	The former is significantly more important than the latter
7	The former is more important than the latter
2, 4, 6	Represents an intermediate value of the above-described adjacency determination

Table 4. Pair of comparative judgment matrix

	G	*V*	*L*
G	1	1/2	1
V	2	1	2
L	1	1/2	1

The judgment matrix J was denoted as

$$J = \begin{bmatrix} 1 & \frac{1}{2} & 1 \\ 2 & 1 & 2 \\ 1 & \frac{1}{2} & 1 \end{bmatrix} \tag{5}$$

The product of each line element of the judgment matrix and the cubic root was calculated.

$$f_1 = (1 \times 1/2 \times 1)^{1/3} = 0.7937$$

$$f_2 = (2 \times 1 \times 2)^{1/3} = 1.5874$$

$$f_3 = (1 \times 1/2 \times 1)^{1/3} = 0.7937$$

Where

$$F = (f_1, f_2, f_3) = (0.7937, 1.5874, 0.7937)^T$$

The vector F was normalized and eigenvector π_{max} was obtained.

$$F = (0.25, 0.5, 0.25)^T$$

$$J * F = \begin{bmatrix} 1 & 1/2 & 1 \\ 2 & 1 & 2 \\ 1 & 1/2 & 1 \end{bmatrix} \begin{bmatrix} 0.25 \\ 0.5 \\ 0.25 \end{bmatrix} = \begin{bmatrix} 0.75 \\ 1.5 \\ 0.75 \end{bmatrix} \tag{6}$$

$\pi_{max} = 1/3 \times (0.75/0.25 + 1.5/0.5 + 0.75/0.25) = 3$

Consistency was tested by CR=CI/RI. Where CR was the ratio of random consistency of the judgment matrix, $CI = (\pi_{max} - n)/(n-1)$ was the general consistency index of the judgment matrix, RI was mean random consistency of the judgment matrix.

For the 1-5th order judgment matrix, the RI values were listed in Table 5. Only when CR <0.10, it can be considered that the judgment matrix was of satisfactory consistency.

Table 5. Average random consistency index of judgment matrix

n	1	2	3	4	5
RI	0.00	0.00	0.58	0.90	1.12

$$CI = (\pi_{max} - n)/(n-1) = (3-3)/(3-1) = 0$$

$$CR = CI/RI = 0/0.58 = 0$$

CR=0<0.1 indicated that the judgment matrix was of satisfactory consistency. So the weight F was normalized as.

$$F = (F_G, F_V, F_L) = (0.25, 0.5, 0.25)$$

S is Relatively Close: The judgment matrix was shown in Table 6.

Table 6.Pair of comparative judgment matrix

	G	V	L
G	1	1/2	1/2
V	2	1	1
L	1	1	1

SO the weight F was calculated as.

$$F = (F_G, F_V, F_L) = (0.2, 0.4, 0.4)$$

During the movement of the arm, the probability of grabbing each target was denoted by P

$$P = \{P_1, P_2, \cdots\cdots, P_n\} \tag{7}$$

$$P_i = F_G * P_{Gi} + F_V * P_{Vi} + F_L * P_{Li} \tag{8}$$

The probability of the intention expression by each dominant factor in the motion of the user's arm was calculated by the parallel processing method which was adopted in the intention inference algorithm. The probability of each target was calculated according to the weight of the distribution using the formula 7, and the goal of maximum probability which was the target location that the user will reach for. The flow chart of the algorithm of intention inference was shown in Figure.5.

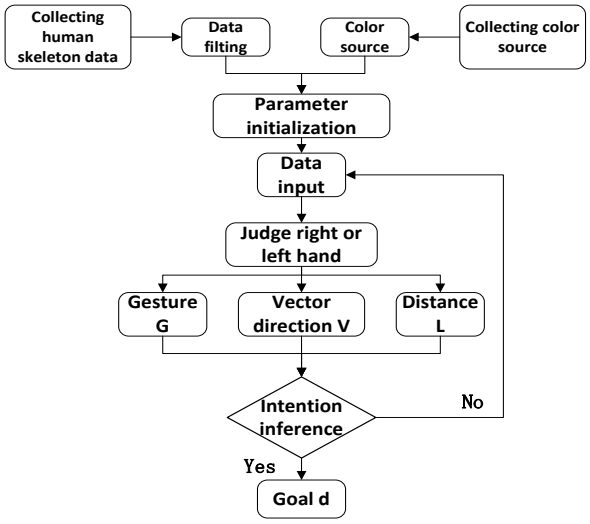


Figure 5. Flow chart of intention recognition algorithm

Experiment Result

The Microsoft's Kinect sensor was used for data acquisition in this paper, combined with Visual Studio 2010+ Kinect SDK +OpenCV to draw up the human skeleton, track information of the bone and display the human bone's movement fully. Due to the lack of accuracy on detection and stability of the Kinect sensor, which is not equal to the professional wearable devices, even if the human body posture is fixed or static, the captured three-dimensional coordinates of the joint point also has small fluctuation. Furthermore, when there's occlusion between the joints, the distortion of captured joint points arises. These factors are not conducive to the control of the robot arm. Therefore, the filter processing should be performed. And the spatial coordinates, velocity, distance from the target and hand gestures were obtained based on the collected data with local Kalman filtering. The OpenCV image processing function was used to recognize the characteristic value of the object on the desktop, including the number, the shape and the spatial coordinates of the object. And the studied object was modeled as the target position which the arm was about to reach. Three series of experiments were designed in this paper. And several experiments were carried out to verify the algorithm respectively. The sampling time was discrete.30 frame data was collected per second.

Experiment 1: Experiment 1 was divided into two groups: (1) Three targets on the experimental platform were used as target object to initialize the parameter. And the algorithm was verified in the absence of filtering. (2) Three objects on the experimental platform were used as target objects to initialize the parameters. And the algorithm was verified. The process of grabbing the target by the user was shown in Figure.6. The metabolic probability of three targets in the process of motion was shown in Figure.7. The change of the coordinate position of the hand during the movement was shown in Figure.8. The metabolic probability of the movement before and after the filtering was shown in Figure.9. The comparative analysis of the experimental result of object prediction before and after filtering was shown in Figure.10.

(a) (b) (c) (d)

Figure 6. The Process of grabbing objects

The process of the user's grabbing and the prediction was shown in Figure.8. According to the position and the identifier of the hand, the target which the user reached for can be predicted by the algorithm in advance and indicated by the green identifier.

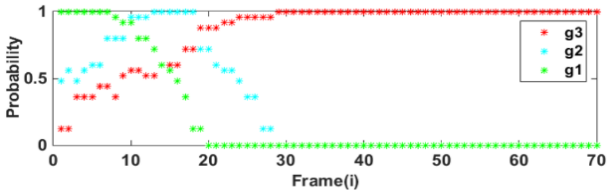
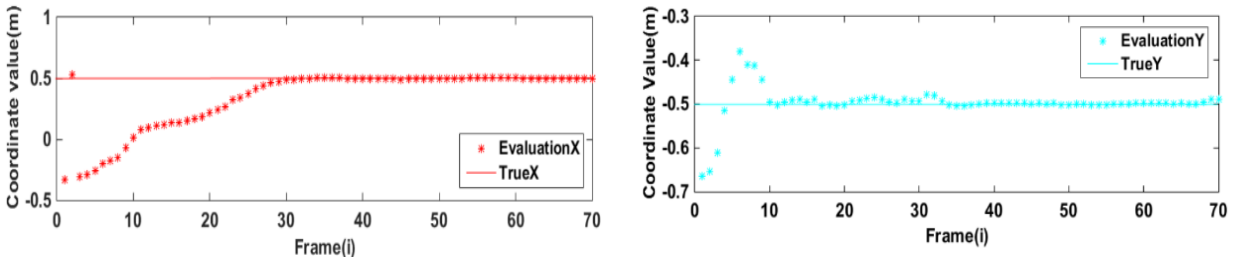


Figure 7. The probability of three goals change

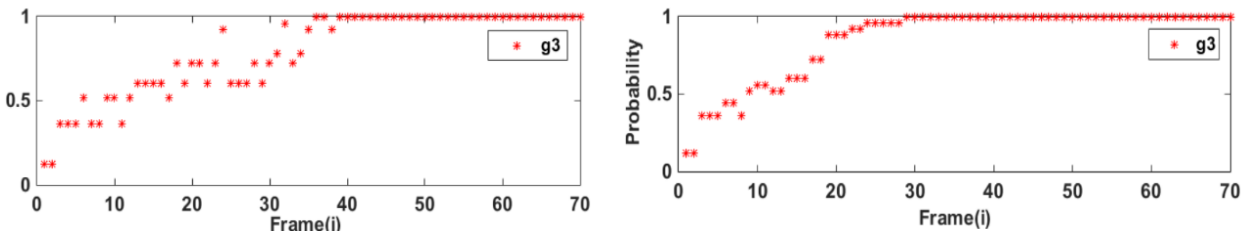
As showed in Figure.7 and Figure.10, the probability of each possible target was calculated using the formula 7 according to the weight of the distribution on the basis of V , G and L at the beginning of grabbing, and the probability of d_1 was the maximum, then d_1 was inferred as the destination. During the movement, the probability of d_2 firstly increased to the maximum ahead of d_3 when the direction and distance of the hand grab was changed, then d_2 was inferred as the destination. In the process of further movement, the probability of d_3 gradually increased to the maximum and came to a stable state, the probability of d_1 and d_2 gradually reduced to zero gradually, then d_3 was inferred as the destination ultimately.



(a) (b)

Figure 8. The position of the hands of evaluation value and real value in the process of movement.

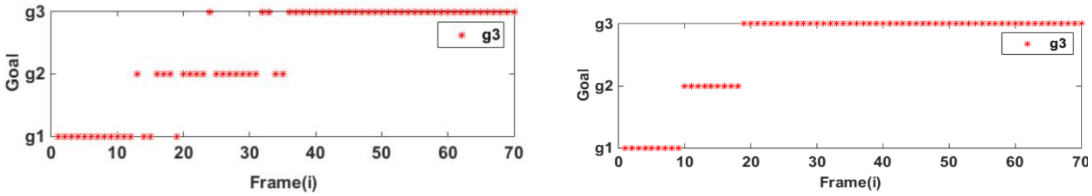
The change of the coordinate position and the position of the target in the process of grasping was shown in Figure.8. It was displayed that the hand reached the target position in the 30th frame. The algorithm can correctly predict the user's intention in the 20th frame according to Figure.7 and Figure.9. The verification algorithm can predict the target position that the user is going to arrive according to the user's hand gesture, hand grasping direction and the distance of the hand position in advance, which shows the algorithm is feasible.



(a) (b)

Figure 9. The probability of before (a) and after (b) changes

Filtering of human skeletal data can eliminate the effect of jitter on probability, improve the accuracy of prediction probability, and reduce the time required to achieve stability in Figure.9.

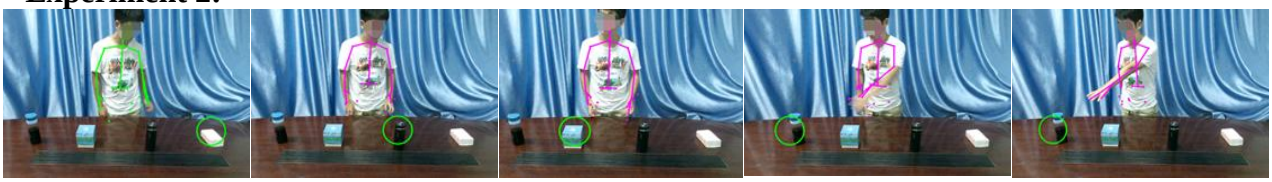


(a) (b)

Figure 10. Predictive results of $d_3(g_3)$ before (a) and after (b) filtering

Filtering can improve the prediction accuracy of the user's intentions, and shorten the time of the advanced prediction in Figure.10.

Experiment 2:



(a) (b) (c) (d) (e)

Figure 11. The process by different users grab the targets

As shown in Figure.11 and Figure.12, four targets on the desktop were used as objects and different users to verify the algorithm.

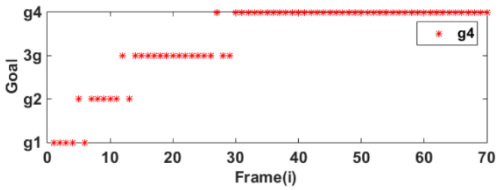


Figure 12. Different user intention inference results

According to the results of experiment 1and 2, the algorithm can be used for different users to accurately predict the user's intention. It was proved that the algorithm was universal, independent of the number of targets and individual users, and could infer the user's intention accurately and effectively.

Conclusion

The destination where the user would reach for was predicted in this paper. Weight matching was performed by four main ways of expressing user's intention to predict the user's intentions. The user's intentions can be predicted quickly and accurately under the experimental conditions of different targets and users through multiple experiments, which fully verified the feasibility of the algorithm. The efficiency of the Human-Computer Cooperation can be improved by the algorithm of intention inference efficiently.

References

1. Bicchi, A., Bavaro, M., Boccadamo, G., & De Carli, D. (2008). Physical human-robot interaction: Dependability, safety, and performance. IEEE International Workshop on Advanced Motion Control (pp.9 - 14).

2. Zhu, C., Sun, W., & Sheng, W. (2008). Wearable sensors based human intention recognition in smart assisted living systems. International Conference on Information and Automation (pp.954-959).

3. Elfring J, Molengraft R V D, Steinbuch M. Learning intentions for improved human motion prediction[J]. Robotics & Autonomous Systems, 2014, 62(4):591-602.

4. Tahboub, K. A. (2005). Compliant Human-Robot Cooperation Based on Intention Recognition. Intelligent Control, 2005. Proceedings of the 2005 IEEE International Symposium on, Mediterrean Conference on Control and Automation (pp.1417-1422). IEEE.

5. Wang, Z., Lling, K., Deisenroth, M. P., Ben Amor, H., Vogt, D., & Sch&#, et al. (2013). Probabilistic movement modeling for intention inference in human-robot interaction. International Journal of Robotics Research, 32(7), 841-858.

6. Goodwin, G. C., & Aguero, J. C. (2005). Approximate EM Algorithms for Parameter and State Estimation in Nonlinear Stochastic Models. Decision and Control, 2005 and 2005 European Control Conference. Cdc-Ecc '05. IEEE Conference on (pp.368-373). IEEE.

7. Morato, C., Kaipa, K. N., Zhao, B., & Gupta, S. K. (2014). Toward safe human robot collaboration by using multiple kinects based real-time human tracking. Journal of Computing & Information Science in Engineering, 14(1), 98-106.

8. Wang, L., & Sun, Z. (2005). Quadratic estimate of opponent's intentions in multi-robot systems based on current behavior. Journal of Tsinghua University.

9. Lasota, P. A., Rossano, G. F., & Shah, J. A. (2014). Toward safe close-proximity human-robot interaction with standard industrial robots. IEEE International Conference on Automation Science and Engineering (pp.339-344).

10. Ravichandar, H. C., & Dani, A. (2015). Human intention inference through interacting multiple model filtering. IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems.