Proceedings of 2018 IEEE 8th Annual International Conference
on CYBER Technology in Automation, Control, and Intelligent Systems
July 19-23, 2018, Tianjin, China

Real-time Human Motion Estimation for Human Robot Collaboration

Jie Kang[1,2], Kai Jia[1,3], Fang Xu[1,3], Fengshan Zou[1,3], Yanan Zhang[1,2] and Hengle Ren[1,2]
1. State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China
2. University of Chinese Academy of Sciences, Beijing 100049, China
3.Shenyang SIASUN Robot & Automation Co., LTD., China, Shenyang 110168, China
{kangjie & jiakai & xufang & zoufengshan & zhangyanan & renhengle}@sia.cn

***Abstract*— In the process of human robot collaboration, safety is of vital importance, especially when the workspaces of human and robot are intersected, and collisions between them should be avoided. To avoid collision accurately, the motion of people must be in charge in real time, and making a reasonable estimate of human motion, so that the robots can make decisions accordingly, and plan their own motion quickly. This paper presents a framework of real-time motion estimation based on human posture which is based on ROS, firstly, the position of human joints is collected through the Kinect, then the gaussian mixture model (GMM) algorithm and EM algorithm are used to cluster and estimate based on the collected coordinate points, and adding labels to each category, which can help get the sequence of the joint, and realize the function of motion estimation. To guarantee the safety of people, this paper also discusses the motion estimation method of human motion trajectory mutation, which avoids the collision in case of emergency. Finally, the experimental results show that the presented framework of real-time motion estimation can describe the human body's movement accurately and make an accurate prediction, not only ensuring the human security, and it's of great significance in improving the production efficiency.**

Keywords—human robot collaboration; GMM; EM; real-time motion estimation; minimum jerk

I. Introduction

The traditional industrial robot which can follow a preprogrammed program is a multi-joint manipulator or multi-degree of freedom robot facing the industrial field. As the application of industrial robots is more and more extensive, there are more and more situations in which people and robots work closely together. In addition, Due to the high loss of traditional robot deployment, and it is unable to meet the needs of small and medium-sized enterprises, as well as the emerging collaborative market, the emergence and development of cooperative robots have been promoted.

Under the background of close collaboration between human and robot, the safety of human becomes the primary issue [2]. Most industrial robots are now isolated from humans, obviously this is not accord with the demand of human robot collaboration (HRC), to guarantee the safety, the cooperative robot needs to keep running at low speed and can detect and react quickly to the collision, at the same time, it can avoid human body timely through dynamic motion planning and monitor human behavior in real time.

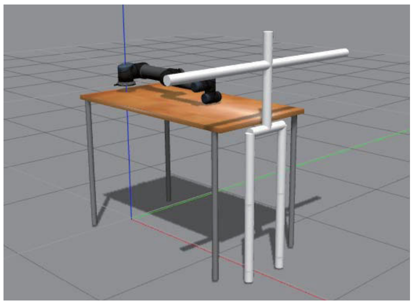

Figure 1. Human robot collaboration platform based on gazebo.

When the robot is avoiding the collision, a path is determined according to the posture information of the current point and target point [3]. The path ensures that the robot will avoid the nearby obstacles and reach the target state without collision. The function of collision avoidance is an important index of robot intelligence, and it is also an important security guarantee for HRC. At the same time, the ability of the robot to actively avoid people or other obstacles in course of operation is also one of the important contents of the research of human robot collaboration. Therefore, the research of robot dynamic collision avoidance is of great practical significance to the promotion and development of HRC.

The paper structure is as follows: the second part introduces the research work related to this thesis, the third part introduces the real-time motion estimation framework proposed in this paper, the fourth part introduces the implementation process of this framework, in the fifth part, experimental results are obtained and analyzed, the sixth part draws the conclusion and looks forward to the future.

II. Related Work

In recent years, whether in the factory or in the family, human robot collaboration is becoming more and more important, to ensure safety and efficiency of HRC and understand the rules of human movement, it is necessary to make the robot adjust its posture and motion path in real time.

The research on human motion estimation can be divided into two aspects: human motion estimation based on moving target and motion estimation based on motion characteristics, the former is to estimate the target point or task to be completed, and then estimate the trajectory of the human body, the latter is to estimate the motion trajectory of the human

978-1-5386-7057-6/18/$31.00 © 2018 IEEE

body directly according to the movement characteristics of people.

J. Mainprice and D. Berenson presents a motion estimation method based on moving target, the gaussian mixture model [4] (GMM) is used to estimate the target location, and then the gaussian mixture regression (GMR) is used to generate the motion trajectory. Elfring, J. Elfring, R. van de Molengraft and M. Steinbuch propose time series analysis [5], the multivariate gaussian distribution is used to represent each time node of the time series.

Song et al. [6] proposed information entropy through measuring individual motion, the dynamic motion of human is given quantitatively, it has high predictability. Based on this work, Pan et al. [7] proposed a linear predictor based on the multivariate normal distribution, however, this method has certain delay and is not conducive to real-time motion planning.

Qiao puts forward a human motion estimation algorithm based on hidden markov model [8], the hidden state and observation state are extracted from a lot of motion data, and according to different types of motion, it predicts the best motion adaptively.

M. Kuderer, H. Kretzschmar, C. Sprunk and W. Burgard propose a method based on maximum entropy principle and consider some characteristics [9], such as the time needed to reach the target position, acceleration, speed of movement and collision avoidance, then finding out the characteristics that can predict the motion of the human very well.

In practical application, motion estimation is a process with high requirement of real time performance, this paper simplifies the model and puts forward the GMM based on posture, which combines minimum-jerk for a hazardous situation, the accurate and efficient estimation is achieved.

III. Real-time Behavior Estimation Framework

The real-time motion estimation framework this paper proposed consists two parts, the first part is estimated under specific task to complete real-time estimation with knowing moving target, the second part is that when human motion changes suddenly, it can still complete the motion estimation more accurately with target motion unknown.

A. Known Moving Target

A GMM needs to be trained in the off-line phase for a specific target job, the training data is collected by Microsoft's Kinect XBOX 360 and converted into a standard data format, as shown in table 1. Human and robot share a workspace and work together, when the hand moves from one known fixed position to another, the robot can quickly identify people's behavior and predict the trajectory to plan their own motion and ensure safety.

TABLE I. THE DATA FORMAT

Number	1	2	3	4
Name	Neck	RShoulder	RElbow	RHand

To have an accurate knowledge of human behavior and early estimate, this paper trains a GMM through human behavior data and classifies people's posture. According to bayesian information criterion (BIC), the number of categories is the number of discrete states when human reaches the target position [11], the more states, the more accurate the behavior estimation, however, the amount of calculation increases and the time cost increases correspondingly, the number of states should not be too small to affect the accuracy of estimation. This paper adds a label to all the above states, specifying the order between each state, thus. In this way, in the practical application process, the corresponding state of the current posture can be clearly defined, and the future state can be obtained quickly, so the real-time performance is strong.

B. Trajectory Sudden Change

In the process of collaboration between humans and robots, especially under long working hours, it is easy for people to be distracted or to be out of work, in this case, it is necessary to ensure the safety of human. However, the gaussian mixture model, which has been trained, cannot meet the requirements at this time, and estimate of human trajectories can vary widely.

For the above considerations, in this paper, on the basis of the original system, the real-time motion estimation method based on the minimum jerk is integrated, this method regards the human's motion trajectory as a smooth motion curve, considers the human body posture as a multi-dimensional vector, and constructs the loss function of the jerk [12], the minimum value of the loss function is obtained by using the variational method and integration by parts, the result is five times polynomial function about the time, so the problem can be transformed to solving coefficients of polynomial, in the process of HRC, depth camera can collect real-time behavior sequence of the human arm, according to the trajectory sequence, using the least square method to get the coefficients of required, thus, the human motion is estimated in real time.

IV. Implementation Process

According to the real-time motion estimation framework, this paper can be divided into three steps to implement. Firstly, clustering algorithm is used to cluster the data set, and then the clustering results are processed to obtain the ideal trajectory, at last, the trajectory change in the case of emergency is dealt with specially.

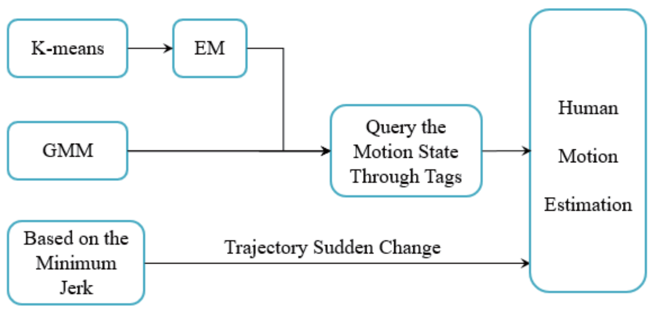

Figure 2. Real-time motion estimation framework.

A. Clustering by Gaussian Mixture Model

Gaussian mixture model (GMM) is a linear combination of the multiple gaussian distribution function, in theory, GMM can fit any type of distribution.

There are many ways of expressing body posture, such as joint coordinates and joint Angle, etc., this paper uses Kinect to collect information of human and can directly obtain the position of joints, so choosing joint coordinates as the representation of the body posture. Data set is represented by matrix M, each of these lines represents a pose of the arm, the number of lines representing the number of pose, denoted by N, the column represents the three-dimensional coordinates of each joint of the arm, representing the dimensions of each pose, represented by D, the matrix contains the coordinates of the four joints of the arm, and the probability density of any posture can be represented by GMM [10].

$$p(\mathbf{x}(t)) = \sum_{k=1}^{N} \pi_k p_k(\mathbf{x}(t)) \quad (1)$$

Where, π_k represents the probability of the k'th gaussian component being selected, $p_k(\mathbf{x}(t))$ denotes the probability density of the k'th gaussian element and is denoted by the following formula

$$p_k(\mathbf{x}(t)) = \frac{1}{(2\pi)^{d/2} |\sum_k|^{1/2}} \exp(-\frac{1}{2}(\mathbf{x}(t)-\boldsymbol{\mu}_k)^T \sum_k^{-1}(\mathbf{x}(t)-\boldsymbol{\mu}_k)) \quad (2)$$

$\sum_k$ represents the positive definite covariance matrix.

The important step in training GMM is to estimate the parameters in the probability density function, for the purposes of this paper, the situation of clustering is unknown and cannot be solved with a simple maximum likelihood estimation. In this paper, the general EM algorithm is used to estimate GMM parameters, which is mainly divided into E and M steps, Step E is to calculate the logarithmic likelihood expectation of the data, Step M iterates the parameter values through the maximization principle, the parameter of new model is expressed as

$$\pi_k = \frac{N_k}{N} \quad (3)$$

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{\mathbf{x}(t) \in S} \mathbf{x}(t) \quad (4)$$

$$\sum_k = \frac{1}{N_k} \sum_{\mathbf{x}(t) \in S} (\mathbf{x}(t) - \boldsymbol{\mu}_k)^2 \quad (5)$$

N_k represents the number of posture of the k'th gaussian component of GMM, the whole dataset is represented by S.

The GMM model is obtained after training until the above parameters converge. To improve the iterative efficiency of EM algorithm, this paper uses k-means clustering algorithm to initialize and estimate the initial parameters.

B. Query the Motion State Through Tags

After the GMM training phase, each sample has its own category. In the process of collaboration, the robot needs to judge the current working area and working trend of human, in the case, the work area of the body is called the obstacle area, to make the robot know the position and change of obstacle area accurately, the paper adopts the mean of the position coordinates $\mathbf{x}'_k$ to obtain the location of the obstacle areas, and can be expressed as

$$\mathbf{x}'_k = \frac{1}{N_k} \sum_{n=1}^{N_k} \mathbf{x}_k(t) \quad (6)$$

$\mathbf{x}_k(t)$ represents the samples of the k'th class, N_k represents the total number of samples belonging to the k'th class.

We need to know the future as well as the current state, this is also the purpose of the paper, so a clear understanding of the sequence of categories is needed, the paper implements this function by establishing tags, numbering from beginning to end. Thus, in the case of knowing the current state, it can quickly estimate all states of the future.

Then, the real-time acquisition of joint coordinates is input into the GMM trained in (A), the current state belongs to the class with the highest probability, after reading the label of this class, the state information can be obtained, which achieves the purpose of real-time motion estimation. An obvious advantage of this method is that we only need to know which class the current state belongs to in the motion estimation, regardless of the previous state, it greatly reduces the computation and improves the real time.

C. Behavior Estimation Based on the Minimum Jerk in the Case of Trajectory Sudden Change

Neville Hogan points out that smoothing can be quantified as a function of jerk which is the third derivative of position. For a person's central nervous system, move the hand or other end-effector smoothly from one point to another, it should minimize the sum of squares along its trajectory [14]. For a special trajectory starting with time $t = t_s$ and ending with time $t = t_f$, the smoothness can be measured by calculating the loss function of jerk.

$$G(x(t)) = \frac{1}{2} \int_{t_s}^{t_f} \dddot{x}(t)^2 \, dt \quad (7)$$

In the type, $x(t)$ denotes displacement and is a function of time t, $\dddot{x}(t)$ denotes jerk, $G(x(t))$ denotes loss function, loss function $G(x(t))$ and displacement $x(t)$ are all scalar.

In this paper, the posture of the arm is a multi-dimensional vector $\mathbf{x}(t)$, which is also a function of time t, which can be expressed as

$$\mathbf{x}(t) = f(t), \quad t_s \le t < t_f \quad (8)$$

Now the loss function $\mathbf{G}(\mathbf{x}(t))$ should be expressed as

$$\mathbf{G}(\mathbf{x}(t))=\frac{1}{2}\int_{t_s}^{t_f}(\ddot{\mathbf{x}}(t))^T\ddot{\mathbf{x}}(t)dt \tag{9}$$

To find the minimum value of this function, Hogan used a technique called variational method. The idea is similar to finding the minimum value of a function: finding the derivative of the function with respect to a small disturbance, a minimum is found when the derivative is zero.

Assumed that the starting and ending velocities and accelerations are zero [15] in the point to point movement, the hypothesis basically meets the actual demand. If an accessory is moved from one position $\mathbf{x}=\mathbf{x}_s$ to the end $\mathbf{x}=\mathbf{x}_f$ by human's hand in $t=d$ second, the trajectory that satisfies the minimum jerk is

$$\mathbf{x}(t)=\mathbf{x}_s+(\mathbf{x}_f-\mathbf{x}_s)(10(\frac{t}{d})^3-15(\frac{t}{d})^4+6(\frac{t}{d})^5) \tag{10}$$

As you can see, the solution of optimization problem is a quintic polynomial, namely the motion trajectory of the human arm can be fitted with a five-degree polynomial, to describe the motion of the arm more accurately, the problem of real time motion estimation is transformed into the problem of a five-time polynomial curve fitting based on least square method, so that the polynomial coefficient can be applied to the actual situation more accurately.

The Kinect is used to collect arm posture in real time and obtain a series of behavioral sequences $\mathbf{x}(t)$, the five polynomial coefficients to be fitted are represented by vector $\mathbf{q}$, the estimator $\hat{\mathbf{x}}(t)$ can be expressed as

$$\hat{\mathbf{x}}(\mathbf{q})=g(t,\mathbf{q}) \tag{11}$$

According to the basic principle of least square method, the loss function of polynomial coefficient vector can be obtained.

$$H(\mathbf{q})=\frac{1}{2}\sum\left\|\hat{\mathbf{x}}(\mathbf{q})-\mathbf{x}\right\|^2 \tag{12}$$

By minimizing the above loss function, the required coefficient vector can be obtained, thus, real-time motion estimation is obtained. This method also has a good predictive effect when the target of human movement is unknown. The problem that GMM cannot adapt to trajectory mutation is solved, and the environmental adaptability of the system is improved.

V. EXPERIMENT AND ANALYSIS

Based on the theoretical analysis and through a series of experiments and data analysis, we verify the feasibility and effectiveness of the method considering HRC scenarios. Firstly, the number of clustering categories in GMM is analyzed, and the most suitable number of clustering categories is obtained, then evaluating the gaussian mixture model and the error between the estimated trajectory and the actual trajectory is analyzed, finally the motion estimation of change of trajectory or special case is analyzed, so as to fully guarantee the safety in the process of HRC.

A. Clustering

When we train the GMM, a very important problem is the number of cluster categories, it affects the quality of training model and the accuracy of motion estimation, in the process of human robot collaboration, security is the most important, and precision of motion estimate affects the safety directly, so the accuracy of behavior estimation is an important index to determine the number of categories.

The higher the number of categories, the higher the model complexity and the higher the cost, but it's too few to express enough characteristics, according to BIC

$$BIC=-2ln\hat{L}+k\bullet ln(n) \tag{13}$$

Among them, $\hat{L}$ is maximum likelihood estimation [9], k is the number of free parameters for the model, n is the number of collection value. By calculating, the number of categories is about 25, to make the model more precise, we have an error analysis of each case when the number of categories is between 20 and 27 and obtain the sum of error squares between the estimated trajectory and the actual trajectory in each case, as shown in figure 3. The horizontal coordinate represents the sum of error squares and the ordinate represents the number of categories. the vertical coordinate adopts the form of error squares in the analysis and comparison of the model.

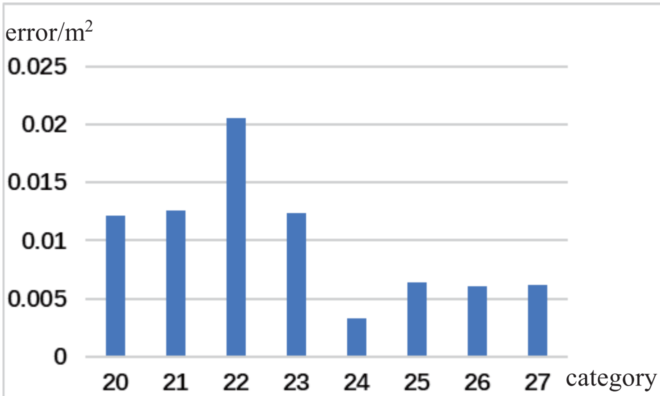

Figure 3. Clustering analysis.

Figure 3 shows the estimate error when choosing different number of category. It is clear from the figure that the sum of error squares is minimized when the data set is divided into 24 classes and is less than 0.005, it's basically going to coincide with the actual trajectory and the error can be negligible. In other cases, the sum of error squares is greater than 0.005 or higher. It also reflects that the GMM is suitable for motion estimation.

B. Real-time Behavior Estimation

The scenario in this paper is in the process of HRC, the operator's hand moves from the starting point to the target point. The data set collected by Kinect is divided into 24 categories, GMM is trained to obtain the weight of each category which is the weight of each component in GMM. Figure 4 describes the comparison of weights, so you can see the number of samples contained in each category.

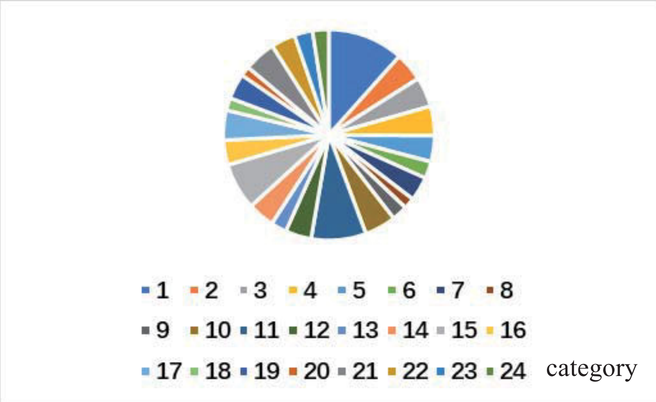

Figure 4. Weight of each category.

The first 1000 poses were selected from the collected data set as the training data set, using EM algorithm to estimate the model parameters, after that each pose will have its own category. We add labels to each pose based on the number of category, so in the process of estimation, you can find the corresponding label, according to each category tag and effectively estimate the posture of the future.

After adding the label, the error between the neck position and the actual position is analyzed, because people work across the robot and share the same workspace, position of neck changes little, the error analysis as shown in figure 5.

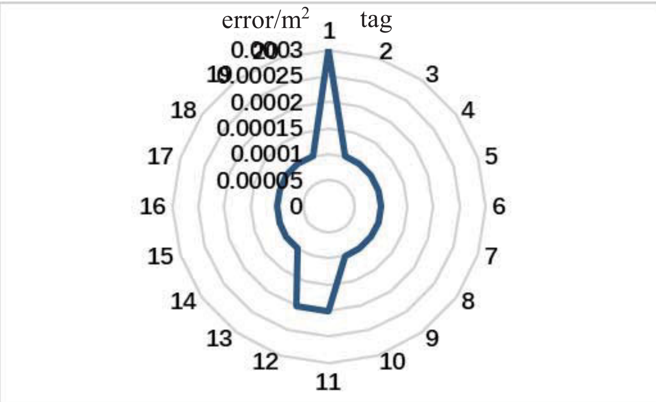

Figure 5. The squared error of the cervical coordinates.

Figure 5 indicates that the error of neck is very small, the estimated position and actual position almost in the same place, the sum of error squares is mainly concentrated in around 0.0001, the error of spatial position is about 0.01m. In other words, the sum of error squares in the three - dimensional coordinates of the neck can be ignored which is no more than 0.0003, the validity of gaussian mixture model to real - time motion estimation is proved, which indicates that it can be used to solve such problems proposed by the paper.

In the human assembly operation, the arm end, namely the hand, has a large range of changes and high uncertainty. The error analysis of the motion of the arm is very representative, which reflects the ability of the model to solve this problem, from this perspective, we compare two models, the first is based on the minimum-jerk, which is under the circumstances of less samples, the changes of error is given in figure 6.

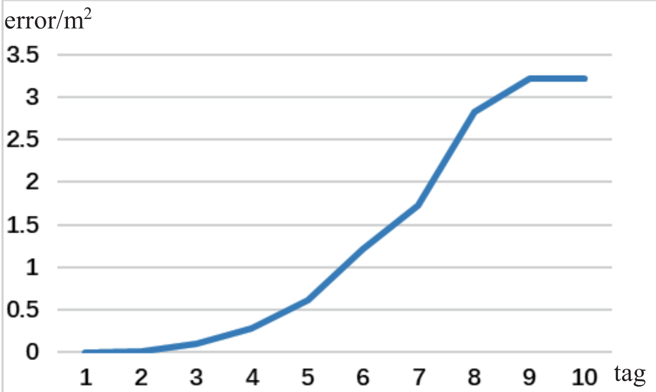

Figure 6. Error change based on minimum jerk.

Figure 6 shows that the error of motion estimation becomes larger and larger when human arm moves, it eventually will deviate from the actual state and become more and more distant, and the reliability is low. The second method adopts GMM, we also analyze changes of error and obtain figure 7.

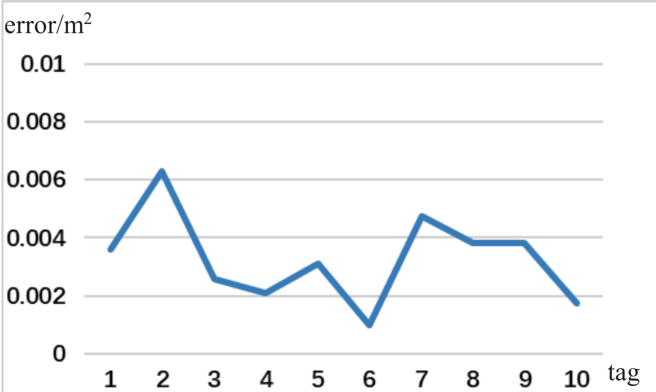

Figure 7. Error change based on GMM.

Due to the error of the second method is smaller than that of the first one, and the difference is large, it is not convenient to display in same graph. But by comparing figure 6 and figure 7, method of GMM not only has smaller error obviously, and the margin of error is small, the fluctuation is basically in the range of 0.001 and 0.006, which is relatively stable. On the other hand, GMM is trained offline before and well describes the task of assignment, it can quickly get the current state and quickly draw future states in real time motion estimation, it greatly reduces the amount of computation and saves time.

C. Real-time behavior estimation when trajectory changes

In HRC, people are easily distracted or out of work, which greatly affects people's safety, the paper fully considers this, a real-time motion estimation method based on minimum-jerk is proposed to solve this problem, such problem is characterized by strong suddenness, unclear target state and uncertain behavior. Therefore, it is required that the real time is strong and the change of trajectory can be estimated accurately.

The initial error of the figure 6 is very small, but as time goes on, the error gets bigger and bigger. Figure 8 depicts that the error of estimation decreases rapidly while collecting data is continuously enriched, therefore, the data can be collected in the initial stage frequently, and the error can be effectively reduced or even ignored.

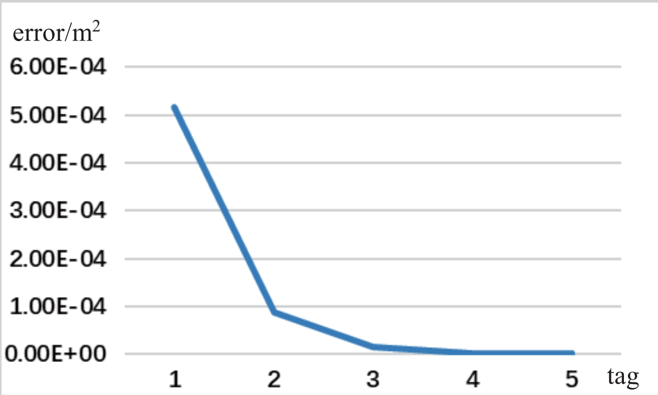

Figure 8. The error varies with the sample size.

Based on the minimum-jerk, Hogan deduced that the trajectory of human motion can be fitted by five polynomial, to show the behavior trajectory and situation of fitting more clearly, the figure 9 shows the projection of the fitting curve on the XOY plane with sufficient samples and exchanges the actual horizontal axis and vertical axis.

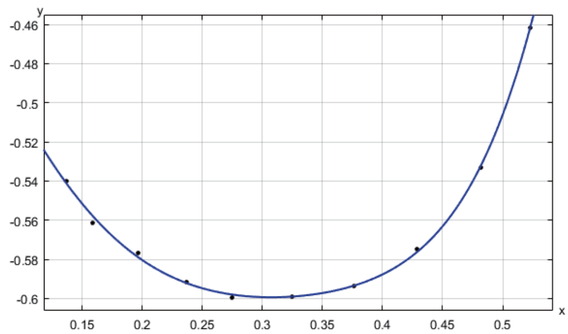

Figure 9. Five polynomial fitting curves.

As you can see from figure 9, the five-degree polynomial perfectly describes the human behavior trajectory and almost completely overlaps with the actual trajectory coordinate points, the standard deviation is only 0.002543, coefficient of determination R^2 is 0.9984, which is proved that the algorithm is highly predictive for human motion trajectory.

VI. Conclusion and Future Work

In the paper, a new real-time motion estimation framework is proposed, which fully guarantees the safety and minimizes the hidden danger in the process of HRC.

The proposed framework is based on collaborative environment of human and robot, robot can acquire and estimate the change of body posture in real-time. To facilitate the motion planning, framework consists of two parts, the first part is the normal process of collaboration, the robot defines the movement trajectory and moving target, GMM and EM algorithm are used to realize fast and real-time motion estimation, the second part is to deal with emergency in time, the robot does not know human's attitude change and the movement target, adopting the method of the minimum jerk, experiment proves that the method is accurate, efficient and able to handle special events in a timely manner.

In terms of human robot collaboration, the most important thing is safety, any work shall be carried out under the premise of security. There are usually very few things on the workbench, the environment is relatively simple, future work should be focused on the application in specific environment, and finish relatively complex task, realizing efficient production of workshop, it's also where we're going.

Acknowledgment

This work is supported by the National Key Research and Development Program of China under Grant No.2017YFF0107800.

References

[1] Shariatee M, Khosravi H, Fazl-Ersi E. Safe collaboration of humans and SCARA robots[C]// International Conference on Robotics and Mechatronics. IEEE, 2017:589-594.
[2] A. Bicchi, M. Peshkin, and J. Colgate, "Safety for physical humanrobot interaction", in Springer Handbook of Robotics, pp. 1335-1348, B. Sciliano and O.Khatib, Eds. Springer, 2008.
[3] Pellegrinelli S, Admoni H, Javdani S, et al. Human-robot shared workspace collaboration via hindsight optimization[C]// Ieee/rsj International Conference on Intelligent Robots and Systems. IEEE, 2016:831-838.
[4] Stauffer C, Grimson W E L. Adaptive Background Mixture Models for Real-Time Tracking[C]// cvpr. IEEE Computer Society, 1999:2246.
[5] Horn B K P, Schunck B G. Determining optical flow[J]. Artificial Intelligence, 1981, 17(1–3):185-203.
[6] Song CM, Qu ZH, Blumm N, Barabsi AL. Limits of predictability in human mobility. Science, 2010,327(5968):1018−1021.
[7] Pan TL, Sumalee A, Zhong RX, Indra-Payoong N. Short-Term traffic state prediction based on temporal-spatial correlation. IEEE Trans. on Intelligent Transportation Systems, 2013,14(3):1242−1254.
[8] Qiao SJ, Shen DY, Wang XT, Han N, Zhu W. A self-adaptive parameter selection trajectory prediction approach via hidden Markov models. IEEE Trans. on Intelligent Transportation Systems, 2015,16(1): 284−296.
[9] M. Kuderer, H. Kretzschmar, C. Sprunk, and W. Burgard, FeatureBased Prediction of Trajectories for Socially Compliant Navigation,in Proceedings of RSS, (Sydney, Australia), 2012.
[10] J. Mainprice and D. Berenson, Human-robot collaborative manipulation planning using early prediction of human motion, in Proceedings of IROS, pp. 299–306, IEEE, Nov. 2013.
[11] Burke, Michael. (2015). Fast upper body pose estimation for human-robot interaction. 10.17863/CAM.203.
[12] Dinh K H, Oguz O, Huber G, et al. An approach to integrate human motion prediction into local obstacle avoidance in close human-robot collaboration[C]// IEEE International Workshop on Advanced Robotics and ITS Social Impacts. IEEE, 2015:1-6.
[13] C. M. Bishop, Pattern recognition and machine learning (information science and statistics). Springer, 2006.
[14] Shadmehr R, Wise S P. The computational neurobiology of reaching and pointing : a foundation for motor learning[M]. MIT Press, 2005.
[15] Ren O. Prediction of Human Pointing Gesture with Minimum-Jerk Model for Human-Robot Interaction and Its Evaluation[J]. Journal of Japan Society for Fuzzy Theory & Intelligent Informatics, 2017, 28(6):911-919.