01 May 2019

Vision Sensor based Action Recognition for Improving Efficiency and Quality under the Environment of Industry 4.0

Zipeng Wang

Ruwen Qin
Missouri University of Science and Technology, qinr@mst.edu

Jihong Yan

Chaozhong Guo

Recommended Citation

Z. Wang et al., "Vision Sensor based Action Recognition for Improving Efficiency and Quality under the Environment of Industry 4.0," *Procedia CIRP*, vol. 80, pp. 711-716, Elsevier B.V., May 2019.
The definitive version is available at https://doi.org/10.1016/j.procir.2019.01.106

Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 80 (2019) 711–716

www.elsevier.com/locate/procedia

26th CIRP Life Cycle Engineering (LCE) Conference

Vision Sensor Based Action Recognition for Improving Efficiency and Quality Under the Environment of Industry 4.0

Zipeng Wang[a], Ruwen Qin[b,*], Jihong Yan[a], and Chaozhong Guo[a,*]

[a] School of Mechatronics Engineering, Harbin Institute of Technology, Xidazhi 92, 150001 Harbin, China
[b] Department of Engineering Management, Missouri University of Science and Technology, Rolla, 65409 MO USA

* Corresponding author. Tel.: +86-0451-8640-2972 ; fax: +86-0451-8640-2972. *E-mail address:* qinr@mst.edu (R. Qin), cguo@hit.edu.cn (C. Guo).

Abstract

In the environment of industry 4.0, human beings are still an important influencing factor of efficiency and quality which are the core of product life cycle management. Hence, monitoring and analyzing humans' actions are essential. This paper proposes a vision sensor based method to evaluate the accuracy of operators' actions. Each action of operators is recognized in real time by a Convolutional Neural Network (CNN) based classification model in which hierarchical clustering is introduced to minimize the effects of action uncertainty. Warnings are triggered when incorrect actions occur in real time and applications of action analysis of workers on a reducer assembling line show the effectiveness of the proposed method. The research is expected to provide a guidance for operators to correct their actions to reduce the cost of quality defects and improve the efficiency of workforce.

Peer-review under responsibility of the scientific committee of the 26th CIRP Life Cycle Engineering (LCE) Conference.

Keywords: action recognition, convolutional neural network, hierarchical clustering, real-time monitoring

1. Introduction

With the advent of intelligent manufacturing and intelligent workshop, comprehensive innovation has occurred over the past few years to improve the core competitiveness of industry. Product life cycle management is a key method to enhance core competitiveness. In most of the manufacturing scenarios, human beings play an important role in product life cycle management. Hence, it is essential to monitor humans' actions in production by using modern science and technology. Human action recognition technology can provide information about the identity of a person and psychological state accurately, which makes the monitoring and analysis of humans' actions much easier.

During the past few decades, a lot of RGB (R-Red, G-Green, B-Blue) video-based action recognition approaches have been proposed [1]. However, data captured from RGB video is sensitive to the recording environment which always leads to high cost for keeping recording environment. Intuitively speaking, human body can be represented by a series of skeleton joints which can handle the above problem effectively. With the advent of 3D visual sensors such as Kinect, extracting skeletons of human body becomes much easier. Shotton et al. [2] proposed hidden Markov model to quickly estimate human skeleton which has resulted in an interest in skeleton & joint based action recognition. Raviteja Vemulapalli [3] established a rotating 3D spatial model for all parts of the body. Recently, Liu [4] et al. proposed a new method to identify mankind's daily life actions by using Kinect sensor. Huy-Hieu Pham et al. [5] establishes a deep residual network whose recognition accuracy is 99%. However, it hasn't been used in industrial scenarios. Jun Yang et al. [6] use depth images to recognize the actions of construction workers, which is a method with the highest accuracy in the recognition of complex actions in specific scenarios. Although there are many kinds of approaches mentioned above, none of them has been applied to

Peer-review under responsibility of the scientific committee of the 26th CIRP Life Cycle Engineering (LCE) Conference.
10.1016/j.procir.2019.01.106

recognize industrial production actions. In this paper, a real-time action recognition system for assembling line actions based on convolution neural network and hierarchical clustering is proposed along with a specific industrial action dataset to lay a foundation for the research of real-time action recognition in industrial field.

The rest of the paper is organized as follows: Section 2 discusses the industrial actions dataset we established. In Section 3, the details of our proposed method are presented. The real-time recognition function is also discussed in this section. Experiment results are shown in Section 4. Finally, Section 7 concludes the paper and discusses our future work.

2. Establishment of industrial basic action dataset

2.1. Scheduling problem formulation

The availability of appropriate datasets is essential for action recognition of specific scenarios. Considering the current status of action dataset research, there are many kinds of action datasets, e.g., KTH action dataset for mankind's daily life, UCF action dataset for sports actions and so on, but there is no specific dataset of industrial actions. Hence, establishing a specific action dataset is very important for our follow-up work. A typical action database establishment process generally contains two major components, action classification and action samples selection. In action classification, Gilbreths proposed 18 therbligs which were regarded as the most authoritative action classification shown in Fig.1. Unfortunately, some of actions proposed by Gilbreths such as 'plan', 'avoidable delay' cannot be directly observed and recorded. And the definitions of some actions such as 'Use' are too rough to be utilized directly. Hence, in order to cover the possible actions in actual production as comprehensively as possible, subdivision of the proposed therbligs and the addition of new actions are essential.

Therblig	Color	Symbol/Icon	Therblig	Color	Symbol/Icon
Search	Black		Use	Purple	
Find	Gray		Disassemble	Violet, Light	
Select	Light Gray		Inspect	Burnt Orange	
Grasp	Lake Red		Pre-Position	Sky Blue	
*Hold	Gold Ochre		Release Load	Carmine Red	
Transport Loaded	Green		Unavoidable Delay	Yellow Ochre	
Transport Empty	Olive Green		Avoidable Delay	Lemon Yellow	
Position	Blue		Plan	Brown	
Assemble	Violet, Heavy		Rest for overcoming fatigue	Orange	

Fig. 1. 18 therbligs proposed by Gilbreths in 1920

To gain more practical actions of industrial production, data statistics-based method [7] was introduced. In this paper, statistical analysis was applied to 20 production process videos of different industrial production lines with different products. Based on therbligs of Gilbreths and the statistical analysis result, 11 actions shown in Fig.2 were ultimately determined as the basic actions to establish database due to their high occurrence frequency. And in action samples selection, there are two basic principles needed to be followed [8]:

(1) Sample size (the amounts of operators) ought to be guaranteed which is usually between 10-15.

(2) Operators ought to contain different genders (male/female), figures (height/fat or thin), operation habit (left handedness/ right handedness) etc.

Therefore, 15 operators who were required to have different genders, figures and operation habits were employed to record action data for the database and ensure the diversity of action data.

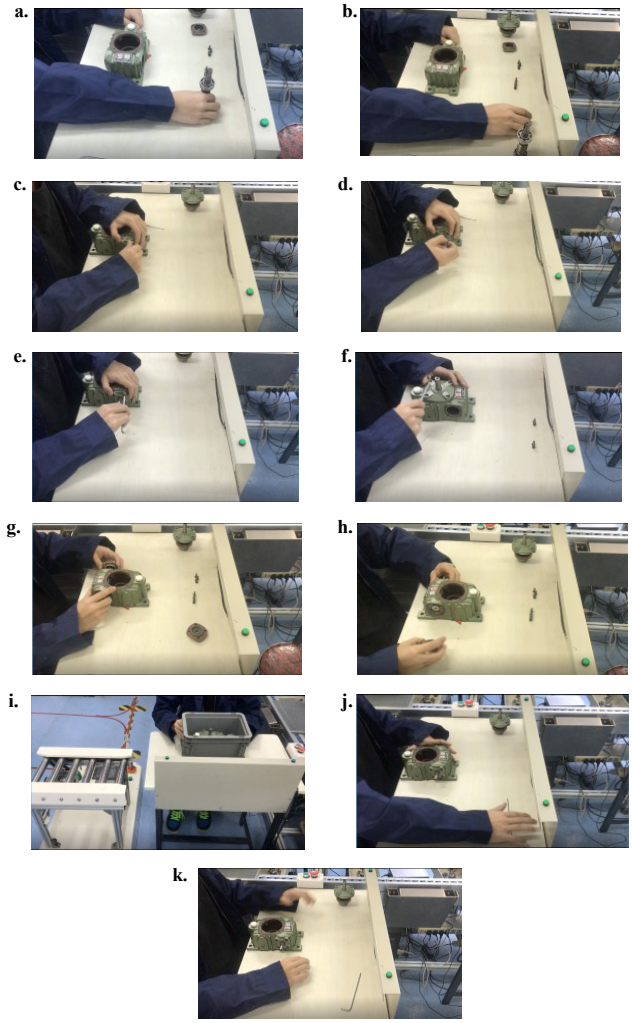

Fig. 2. (a) action 1--- 'grab'; (b) action 2--- 'return'; (c) action 3--- 'insert'; (d) action 4--- 'pullout'; (e) action 5--- 'tighten'; (f) action 6--- 'hit'; (g) action 7--- 'assemble'; (h) 'action 8'---disassemble; (i) 'action 9'--- 'hands handling'; (j) action 10--- 'transport loaded'; (k) action 11--- 'transport empty';

3. CNN based solution for action recognition

3.1. Feature extraction and selection

With the rapid advance of depth-sensing time-of-flight cameras, e.g., Microsoft Kinect sensor or ASUS Xtion, more

detailed and accurate 3D motion structure information can be extracted such as real-time skeleton coordinates. And what's more important is that HAR methods using depth cameras can avoid many obstacles which make HAR a challenging task such as view point, occlusion or lighting conditions. Above all, a skeleton-based method is proposed in this paper.

Existing joint-based action recognition approaches mainly focus on different kinds of angles' calculation method. Gavrila et al. [9] proposed a method of using joints angle to represent human actions. Ofli et al. [10] proposed another approach where many interpretable measures such as the mean of the joint angles were used to represent human actions. Nevertheless, using single feature cannot fully and accurately represent human actions. Furthermore, distance of joints is also a remarkably effective feature which can enrich the detail of actions ignored by single angular eigenvalue. Hence, distance and angle features were combined as multiple recognition features to capture the information of actions more accurately in this paper. Since all the eleven basic assembly actions were upper limb actions, only the upper limb joints were used to calculate the two types of features. There were 17 angle eigenvalues and 17 distance eigenvalues left.

In this paper, Kinect vision sensor shown in Fig. 3(a) which has been applied widely by many researchers was chosen to capture human joint points in real-time. Kinect is capable of capturing 25 joints of human body at the same time, and record data of each joint as the three dimensional coordinates (x, y, z). At the same time, the recording frequency of joint points is set to 30 frames per second to ensure the accuracy and continuity of action recognition.

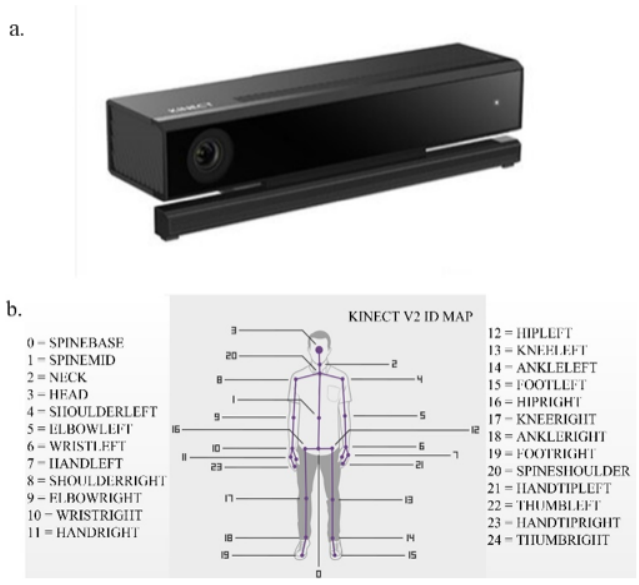

Fig. 3. (a) Kinect: sensor of action recognition; (b) 25 joints of human body

Distance Feature: In the calculation of distance feature, human coccygeal bone was chosen as the origin of coordinates as Fig. 4(a) shown. Assuming that the current spatial coordinate of the origin of the coordinates is (x_0, y_0, z_0), and the current spatial coordinate of *ith* joint point is (x_i, y_i, z_i). The distance feature vector $(d_1, d_2, \ldots, d_{17})$ which contains 17 distance eigenvalues could be calculated by Eq. (1):

$$d_i = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2}\,(i = 1,...,17) \tag{1}$$

Angle Feature: two principles should be followed in angular eigenvalues calculation:

(1) Calculate the angles between adjacent skeletons.

(2) Calculate the angular eigenvalues by the method of vector angle calculation.

Since the value range of vector angle is 0~π, the angle features can be calculated by calculating the cosine value of vector angle. Let vectors $(\overrightarrow{BA}) = (x_1, y_1, z_1)$ and $(\overrightarrow{CB}) = (x_2, y_2, z_2)$ shown in Fig.3(b). According to the Eq. (2) and (3), the angle feature vector $(\theta_1, \theta_2, \ldots, \theta_{17})$ was calculated:

$$\cos\theta_i = \frac{\overrightarrow{BA} \bullet \overrightarrow{CB}}{\left|\overrightarrow{BA}\right| \times \left|\overrightarrow{CB}\right|} = \frac{(x_1 x_2 + y_1 y_2 + z_1 z_2)}{\sqrt{x_1^2 + y_1^2 + z_1^2}\sqrt{x_2^2 + y_2^2 + z_2^2}} \tag{2}$$

$$\theta_i = \arccos\cos\theta_i \tag{3}$$

By synthesizing the two features, the final eigenvector $(\theta_1, \theta_2, \ldots, \theta_{17}, d_1, d_2, \ldots, d_{17})$ for action recognition was obtained. And then the eigenvector was normalized, which can decrease the effect of different ranges of data values on action recognition.

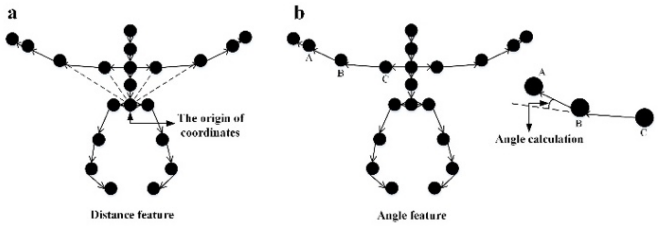

Fig. 4. (a) schematic of distance feature; (b) schematic of angle feature.

3.2. Hierarchical clustering & CNN based action recognition model

The model applied to action recognition is based on convolution neural network (CNN) which has enough reputation in image recognition field. Bilen et al. [11] proposed a dynamic image representation by using raw image pixels of a sequence. However, this approach requires RGB with high resolution. Aiming at solving the problem and improving the stability and accuracy of recognition model, distance & angle features were calculated when the raw data was captured. This process was called data pre-processing. And then, pre-processed data was input into CNN model to extract more effective features automatically which is similar to image recognition process. Intuitively speaking, the principle of image recognition is recognizing the pixel matrix of raw image. According to the feature extraction process mentioned in Section 2 and the habit of human observation, the input

eigenvalue matrix of action recognition model is formed by eigenvalue vectors of all 30 frames as following Eq. (4) shown:

$$\begin{bmatrix} \theta_{1,1} & \cdots & \theta_{1,17} & d_{1,1} & \cdots & d_{1,17} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \theta_{30,1} & \cdots & \theta_{30,17} & d_{30,1} & \cdots & d_{30,17} \end{bmatrix}_{30*34} \quad (4)$$

The output of CNN was a vector shown in Eq. (5). The vector was formed by 1 & 0 in which the position of 1 represented the action label of current action.

$$(0,0,\cdots,1,\cdots 0) \quad (5)$$

By using leave-one-out (LOO) method [12], 6 operators' action data was chosen to test the CNN based action recognition model. Table 1 shows the recognition accuracy.

Table 1: Accuracy of action recognition model without hierarchical clustering

Subject	Training accuracy	Validation accuracy	Testing accuracy
s_1	99.71%	93.36%	27.39%
s_2	99.67%	93.61%	29.87%
s_3	99.51%	90.27%	47.26%
s_4	99.96%	93.21%	56.06%
s_5	99.71%	93.31%	47.03%

Aiming at figuring out the reason of low accuracy, confusion matrix was introduced to the recognition result of each action. By performing action recognition on 15 operators in turn, Fig. 5 shows the confusion matrix of 15 operators. The horizontal coordinates of the confusion matrix represent the prediction action labels and the vertical coordinates represent the actual action labels. It's expected that the number is mainly concentrated on the diagonal line of the confusion matrix which represents no confusion exists. The value outside the diagonal represents the degree of confusion between actions. Taking action 1 shown in Fig. 5 as an example, Action 1 actually has 423 sequences of data, among which 173 sequences are correctly identified and 89 sequences are mistakenly identified as action 2, accounting for 21.04%. It indicates that there is similarity among actions, and the higher the proportion of incorrect recognition is, the greater the similarity is. A hierarchical clustering approach was introduced in order to minimize the effect of action similarity on the accuracy of action recognition. According to confusion matrix, the 11 basic actions were grouped into three layers and seven separate action recognition models (from M1 to M7) shown in Fig. 6 which were trained separately to guarantee that the recognition accuracy of each layer was the highest.

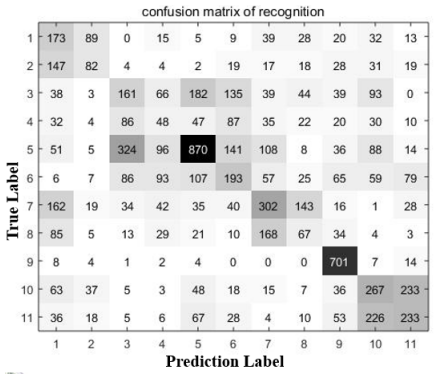

Fig. 5. Confusion matrix for determining degree of confusion.

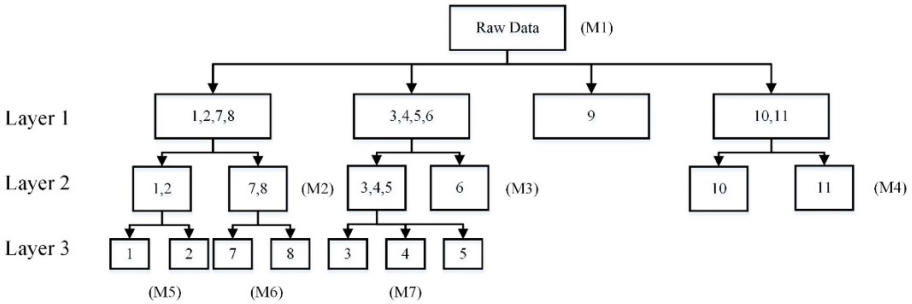

Fig. 6 Block diagram of three-layer hierarchical clustering

For instance, the process of action 'assemble-7' and action 'disassemble-8' may contain the process of action 'grab-1' and action 'return-2'. Therefore, the four actions could be considered as one class in layer 1. However, actions grab/return and actions assemble/disassemble have different purposes. Hence, these two kinds of actions are distinguished in layer 2. And then each action must be specifically classified in layer 3.

3.3. Realization of real time recognition and feedback function

The raw data of each frame captured by Kinect was stored in a sequence of data. Zhu et al. proposed an approach where skeletal data were being generated frame by frame [13]. However, this method with high recognition frequency which could amplify the effect of actions' similarity wasn't suitable for the real-time recognition of such high complexity and high similarity assembling actions. Therefore, an action recognition approach with recognizing frequency of one second was proposed in this paper. The last 30 frames data of current data sequence were got to be the input of CNN model and then the current action label ought to be got as the 'Actual Action' curve shown in Fig. 7. The standard action curve was determined by processing sequence corresponding to the actual production line. The standard action label points didn't coincide with the actual recognition action label points in 9th second and 10th

second which represented that wrong action occurred in Fig. 7 and the alarm was triggered with buzzer to remind operators error occurs at that moment.

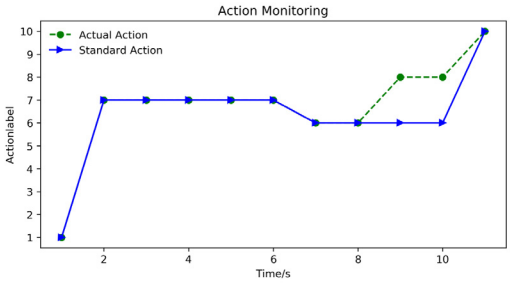

Fig. 7. Profile of real-time action recognition

4. CNN based solution for action recognition

In this section, an experiment was designed to evaluate the proposed hierarchical-clustering based action recognition model.

Equipment setup: (1) Kinect was placed at the distance of 1.5m from the operating platform and 1 m from the ground. (2) The recording frequency of Kinect was set to 30 fps. (3) The recognition frequency was determined as 1 second.

Implementation details: 6 operators who did not participate in database establishment were employed as the testing samples. The 6 operators completed 11 basic actions in turn and actions were recognized in real time. The accuracy of each operator's 7 corresponding models was recorded separately as $a_{i,j,k}(i = 1 \cdots 6; j = 1 \cdots 7)$where i represents ith operator , j represents jth model and k represents kth kind of accuracy: training accuracy, validation accuracy and testing accuracy and take Table 2 for example.

Table 2. Accuracy recording table of model j

Subject	Training	Validation	Testing
S_1	$a_{1,j,1}$	$a_{1,j,2}$	$a_{1,j,3}$
S_2	$a_{2,j,1}$	$a_{2,j,2}$	$a_{2,j,3}$
S_3	$a_{3,j,1}$	$a_{3,j,2}$	$a_{3,j,3}$
S_4	$a_{4,j,1}$	$a_{4,j,2}$	$a_{4,j,3}$
S_5	$a_{5,j,1}$	$a_{5,j,2}$	$a_{5,j,3}$
S_6	$a_{6,j,1}$	$a_{6,j,2}$	$a_{6,j,3}$

According to the raw data in Table 2, the average value and maximum value of each kind of accuracy were calculated respectively as the evaluation criteria for the accuracy of the model as Table 3 shown.

Table 3. Average and maximum value of each case of model j

Subject	Training	Validation	Testing
Average	$(\sum_{i=1}^{6} a_{i,j,1})/6$	$(\sum_{i=1}^{6} a_{i,j,2})/6$	$(\sum_{i=1}^{6} a_{i,j,3})/6$
Max	$\max_{i\in(1,6)}(a_{i.j.1})$	$\max_{i\in(1,6)}(a_{i.j.2})$	$\max_{i\in(1,6)}(a_{i.j.3})$

According to the formulas shown in Table 3, the average and maximum recognition accuracy of all three kinds of accuracy of the 7 models can be calculated as Tables 4 to 10 shown.

Table 4. Accuracy of recognizing actions 1,2,7,8 & 3,4,5,6 & 9 & 10,11 (M1)

Subject	Training	Validation	Testing
Average	99.68%	97.51%	86.35%
Max	99.94%	98.32%	90.74%

Table 5. Accuracy of recognizing actions 1,2 & 7,8 (M2)

Subject	Training	Validation	Testing
Average	99.27%	97.71%	86.49%
Max	99.86%	99.72%	99.29%

Table 6. Accuracy of recognizing actions 3,4,5 & 6 (M3)

Subject	Training	Validation	Testing
Average	99.52%	97.66%	87.54%
Max	100.00%	98.15%	99.49%

Table 7. Accuracy of recognizing actions 10 & 11 (M4)

Subject	Training	Validation	Testing
Average	94.59%	91.51%	52.75%
Max	100%	97.72%	80.47%

Table 8. Accuracy of recognizing actions 1 & 2 (M5)

Subject	Training	Validation	Testing
Average	98.21%	87.94%	57.16%
Max	100.00%	94.24%	76.19%

Table 9. Accuracy of recognizing actions 7 & 8 (M6)

Subject	Training	Validation	Testing
Average	99.87%	94.09%	76.82%
Max	100.00%	97.26%	87.59%

Table 10. Accuracy of recognizing actions 3, 4 & 5 (M7)

Subject	Training	Validation	Testing
Average	99.73%	97.05%	55.29%
Max	100.00%	99.00%	74.45%

By comparing the accuracy of Table 1 and Tables 4 to 10, the action recognition model based on the hierarchical clustering is more accurate than other traditional methods obviously. It can prove that hierarchical clustering has a good effect on improving the accuracy of action recognition considering the similarity among actions. And in terms of current experimental environment, the processor selected in this paper is Intel(R) Core i7-5500U CPU. The average computation cost is about 0.4407511s for each recognition. In this paper, the recognition interval was set as 1 second. Therefore, the current computing time can meet the requirements of real-time recognition.

And then, the system was applied to an assembly line for reducer production to test its practical application effect. The assembly time of single reducer with or without the assistance of the system was recorded separately for ten times. The average and standard deviation of the 10 groups of data calculated are shown in Table 11.

Table 11. Comparison of assembly time with or without system assistance

Subject	Without system (s)	With system (s)
Average	207.1	159.2
Standard deviation	23.5	11.5
variable coefficient	0.11	0.07

The average assembly time under the two assembly conditions are 207.1s and 159.2s respectively, that is, the average assembly time of the reducer is reduced by 23.13% after applying the system for assistance. Obviously, using this system for auxiliary production can improve the production efficiency of enterprises.

5. Conclusions and future work

This study aims at monitoring industrial actions in real time and providing timely guidance by action recognition. An industrial basic actions database containing 11 basic actions was established which can be used wildly in industrial actions study. Comparing with other certain scenario action datasets such as construction workers dataset, WR dataset [14], the industrial basic action dataset proposed in this paper has higher complexity and similarity. And then, the multi-features fusion was applied to represent human actions more fully and accurately. Combining with the features extracted, the hierarchical clustering based CNN model was proposed to reduce the effect of confusions among industrial actions. The accuracy of the-state-of-art method which was designed for construction worker action recognition is nearly 56%. Obviously, the experiment results illustrated that the accuracy of action recognition method proposed in this paper is close to the-state-of-art methodology. Finally, hierarchical clustering based CNN models and action database were applied to establish a system with the function of monitoring actions in real time and providing useful guidance for operators.

The real-time operator action recognizing and correcting system proposed in this paper, not only can effectively improve the production efficiency, but also ensure the quality of products and reduce the quality cost of products. Therefore, the application of this system can effectively improve the management and control ability of enterprises for the product life cycle and enhance the core competitiveness of enterprises.

However, there are still some limitations existing that can be improved in the future. Although the accuracy of action recognition is close to results of the current best method, the recognition accuracy is still high enough. It is not only due to the structure of the algorithm itself, but also because the objects interacting with the operator has not been taken into account. Hence, more experimental tests on the system combined with object recognition will be designed to further improve the recognition capability of the system. Then, the system will be applied in more variable assembly environments to further verify the reliability of the system in the future.

Acknowledgements

This work is funded by NSF-NSFC (Grand No. 51561125002). This work is partially supported by the National Science Foundation (NSF) grant CMMI-1646162 on cyber-physical systems. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

References

[1] J. K. Aggarwal and M. S. Ryoo. Human Activity Analysis: A Review[J]. ACM Computing Surveys, 2011(43): 1-43.
[2] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time Human Pose Recognition in Parts from a Single Depth Image[C]. In CVPR, 2011(1): 1297-1304.
[3] Raviteja Vemulaplli, Felipe Arrate and Rama Chellappa. Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014(82): 588-595.
[4] M. Y. Liu, H. Liu, and C. Chen. Enhanced skeleton visualization for view invariant human action recognition[J]. Pattern Recognition, 2017(68): 346--362.
[5] Huy-Hieu Pham, Louahdi Khoudour, Alain Crouzil, Pablo Zegers, Sergio A. Velastin. Exploiting deep residual networks for human action recognition from skeletal data[J]. Computer Vision and Image Understanding, 2018, 170: 51-56.
[6] Jun Yang, Zhongke Shi, Ziyan Wu. Vision-based action recognition of construction workers using dense trajectories[J]. Advanced Engineering Informatics, 2016, 30(3): 327-336.
[7] Schuldt, C., Laptev, I., Caputo, B.. Recognizing human actions: a local SVM approach[C]. Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, 2004.
[8] http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html
[9] D. M. Gavrila and L. S. Davis. Towards 3-D Model-based Tracking and Recognition of Human Movement: A Multi-view Approach. In International Workshop on Automatic Face and Gesture Recognition, 1995. 3.
[10] Ofli, Ferda, Chaudhry, Rizwan, Kurillo, Gregorij, Vidal, Rene, Bajcsy, Ruzena. Sequence of the Most Informative Joints (SMIJ): A new representation for human skeletal action recognition[C]. Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on, 2012(2): 8-13.
[11] H. Bilen, B. Fernando, E. Gavves, A. Vedaldi, S. Gould. Dynamic image networks for action recognition, in: Proceeding of the Conference on Computer Vision and Pattern Recognition, 2016: 3034-3042.
[12] Alireza Haji Fathaliyan, Xiaoyu Wang, and Veronica J. Santos. Exploiting Three-Dimensional Gaze Tracking for Action Recognition During Bimanual Manipulation to Enhance Human–Robot Collaboration[J]. Frontiers in Robotics and AI, 2018(5): 1-15.
[13] Zhu Guangming, Zhang Liang, Shen Peiyi, Song Juan. An Online Continuous Human Action Recognition Algorithm Based on the Kinect Sensor[J]. Sensors (Basel, Switzerland), 2016, 16(2): 1-18.
[14] Voulodimos, A., Kosmopoulos, D., Vasileiou, G., Sardis, E., Doulamis, A., Anagnostopoulos, V., Lalos, C., Varvarigou, T.. A dataset for workflow recognition in industrial scenes[C]. Image Processing (ICIP), 2011 18th IEEE International Conference on, 2011: 3249-3252.