An Investigation of Applications of Hand Gestures Recognition in Industrial Robots

Qujiang Lei[1], Hongda Zhang[1, 2], Yang Yang[1], Yue He[1], Yang Bai[1], and Shoubin Liu[*2]
[1]Guangzhou Institute of Advanced Technology, Chinese Academy of Sciences
[2]Harbin Institute of Technology (HIT), Shenzhen

***Abstract*—Hand gesture recognition (HGR) is a natural way of Human Machine Interaction and has been applied on different areas. In this paper, we discussed works done in the area of applications of HGR in industrial robots where focus is on the processing steps and techniques in gesture-based Human Robot Interaction (HRI), which can provide useful information for other researchers. We reviewed several related works in the area of HGR based on different methods including sensor based approaches and vision based approaches. After comparing the two types of approaches, we found that 3D vision-based HGR method is a challenging but promising researching area. Then concerning works of implementation of HGR in industrial scenario are discussed in detail. Pattern recognition algorithms that effectively used in HGR like k-means, DTW etc. are briefly introduced as well. Finally, after reviewing some works done in the area of designing gesture set, we proposed several principles in designing industrial hand gesture commands.**

***Index Terms*—Hand gesture recognition, human Robot Interaction, industrial robot, pattern recognition algorithm, gesture set for industry.**

I. Introduction

Traditional industrial robots have proven their performance on the production line after years of evolution. Relatively speaking, it has the advantages of high speed, high precision and good stability. However, the disadvantages of traditional industrial robots are also obvious: high deployment costs, high thresholds for use, low safety factor, and cumbersome programming. With the demand for more flexibility of industrial robots in the market, collaborative robots that are economical, plug-and-play, simple and intuitive programming, high precision and high security are becoming the favorite of manufacturing and retail industry.

While existing methods of human-robot interaction consist of several ways: direct control through computer programming or robot teach pendant, or more intuitive methods such as vocal control or directly dragging and teaching, each method has its limitations. For instance, it is time-consuming to reprogram robot every time for small changes in a producing line, not to mention the expensive costs of hiring professional programmers.

Compared to direct programming, vocal control and directly dragging and teaching may sound like a great improvement of HRI, but the complexity of factory physical environment shouldn't be ignored. On the one hand, the factory environment can be neat and bright, without dangerous chemicals and suitable for human workers. On the other hand, it can be quite the opposite, dusty, insufficient illumination, noisy because of human voices or the honk of vehicles and even full of toxic gas, in which case directly dragging and teaching method cannot be performed and the loud and multifarious sound landscape prevents vocal control approach from working accurately. In fact, it might not work at all.

In order to improve efficiency and reduce worker ergonomic stress and workload and at the same minimize the costs and time wasting of reprogramming, a more effective communication method between human and robot is of urgent requirement.

Since the beginning of human civilization, hand gesture has been involved commonly in the process of communication between human and human as well as between human and animal. Gestures communicate the meaning of statement said by the human being. They come naturally with or without the words to help the receiver to understand the communication. Simple as they seem, they can convey messages as complex as one's feelings and thoughts with different emotions [1]. Despite the uncertainty of hand gesturing meanings, which depend highly on current working context as well as geographical and cultural background, some gestures used by the army, navy, and air force personnel can also communicate very specific information, which gives the possibility of HRI via hand gestures.

Compared with other HRI methods, communication with robots through hand gestures is not only convenient and natural, but also has other superiority. According to the research work done by P. Barattini, without wearable microphones, voice control of the robot can only be possible within 3 meters [2]. In comparison with vocal control, gesture-based robot control system leaves more freedom for workers, while saving troubles of wearing additional devices. It can also be effective in the case when human workers cannot be within the vicinity of their robot partner, such as working in toxic environment. In addition, because of its simplicity, using hand gesture HRI method can minimize the perceptual and physical workload as well as the training costs for workers.

Manuscript received August 27, 2018; revised June 30, 2019.

doi: 10.18178/ijmerr.8.5.729-741

Therefore, gesture recognition has broad applications in human robot interaction.

In this paper, we looked specifically at applications of hand gestures recognition implemented in industrial robot. Our focus is works done in the area of gesture-based human-robot interaction system and its realization. In section Ⅱ, we elaborate several related works in the area of HGR based on different approaches including sensor based approach and vision approach. A comparison of the two approaches is made to provide useful information for other researchers. Then concerning works of implementation of HGR in industrial scenario are discussed in section Ⅲ, where we mainly focus on the modules contained in their HRI system. Pattern recognition algorithms that effectively used in HGR like k-means, DTW etc. are briefly introduced in section Ⅳ. Finally, after reviewing some works done in the area of designing gesture dataset, we propose several principles of designing industrial hand gesture commands in section Ⅴ.

II. Hand Movement Detection Approaches

According to Mitra, in order to comprehend the meaning of gesture, four aspects needed to be concerned, which are spatial information, pathic information, symbolic information and affective information [3]. These information need to be acquired via specific technology approaches. Current HGR techniques can be divided into two main categories in the aspect of data acquisition i.e. the vision based and the sensor based techniques.

Sensor based recognition collects the gesture data by using one or more different types of sensors. These sensors are attached to hand which record to get the position of the hand and then collected data is analyzed for gesture recognition. Vision based techniques uses one or many cameras to capture the hand images and recognize hand gesture through computer vision. Each detecting technology has its pros and cons.

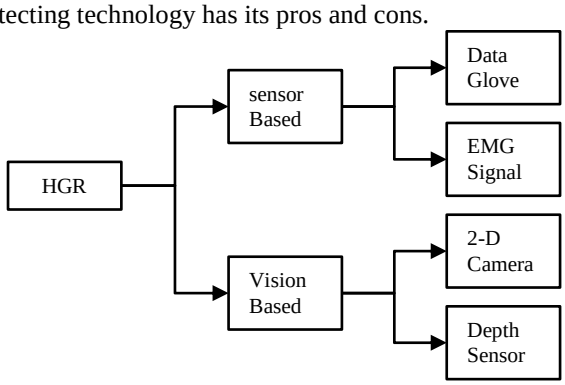

Figure 1. Current approaches of HGR

We will elaborate several related works in the area of HGR based on different approaches (Figure 1.) including sensor based approach (data glove and EMG) and vision approach (2-D image processing based and depth sensor based). A comparison is made for providing information for other researchers.

A. Sensor Based Approach

1) Data glove

Lv et al. design a data glove for gesture recognition [4]. The main processing of their work can be divided into two modules i.e. feature extraction module and gesture classification module. In the first module, they use sparse decomposition model to process the raw data which are transmitted through a serial port by the wireless data glove (Figure 2.) and contain information about the bending angles of each hand joints and the hand positions in Euclidean space.

The result of the decomposition consists of two significant parts: the weight part and the dictionary part which encodes the inherent regular patterns of hand gestures. In the gesture classification module, they feed the dictionary part to an SVM classifier for gesture recognition. Then, when a new data generates, they decompose the data to obtain its dictionary part and send it to the SVM classifier to get its class label. Their gesture recognition system can discriminate four kinds of hand gestures with an accuracy of 93.19%.

Figure 2. The data glove used in Lv's paper [4]

Ge et al. develop a gesture recognition system based on data glove, which can predict incomplete gestures before capturing the entire data in real-time [5]. The systematic framework of their system consists of three important modules i.e. the sensor placement and data collection module, the multimodal fusion and feature extraction module as well as the real-time prediction module (Figure 3.). In the first module, they allocate six flex sensors on the data glove based on the biological characteristics of hand muscle to maintain high precision of data collecting (Figure 4.).

After arranged the flex sensor properly, they build a flex-gesture dataset consists of sixteen hand gestures. Each gesture class contains 3000 six-dimension flex data and a corresponding label for the purpose of training or testing machine learning algorithms. In the multimodal fusion and feature extraction step, they calculate the velocity and the acceleration of each finger and use the moving average method to smooth the velocity. Then they propose a double-finger feature to describe the relationship of adjacent fingertips.

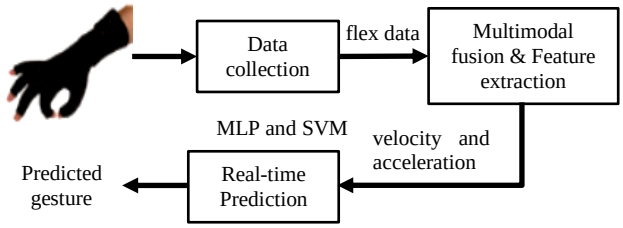

Figure 3. The systematic framework [5]

After these procedures, the raw dataset which contains six-dimension is broadened into n-dimension data with more feature channels: P (position) channel with 6 dimensions, PV (position and velocity) channel with 18 channels as well as PVA (position, velocity and acceleration) channel. If the adjacent finger information is taken into account, the dimension of feature vector doubled. The real-time prediction module mainly implemented through a combination of a multilayer perceptron (MLP) neural network and a multiclass support vector machine (SVM). After trained the models with different kind of feature vector and different machine learning algorithm, they acquire a gesture prediction accuracy of 98.29% with VP feature vector.

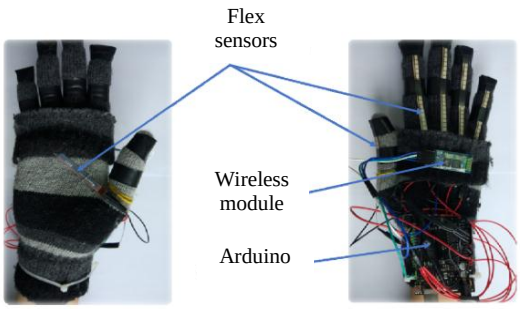

Figure 4. The physical map of data-glove [5]

2) Electromygraphy (EMG) Signal

Ahsan et al. develop a HGR method based on electromygraphy signal and artificial neural network (ANN) [6]. According to Ahsan, electromyography signal is a measure of muscles' electrical activity and usually represented as a function of time, defined in terms of amplitude, frequency and phase. In their gesture recognition system, they use Biopac MP100 data acquisition system to collect EMG signals that generated by moving hands in four directions. In order to extract useful features from the raw EMG data and eliminate all kinds of noise such as electrode noise, motion artifacts, power line noise and inherent noise in electrical & electronic devices, the raw signal are passed through a six order Butterworth band-pass filter and a notch filter and denoised using wavelet method.

The features of EMG signals are extracted from the raw signal via different algorithms listed as follow: MAV (mean absolute value) value, RMS (root mean square) value, VAR (variance) algorithm, SD (standard deviation) algorithm, ZC (zero crossing) algorithm, WL (waveform length) algorithm and SSC (slope sign change) algorithm. For the purpose of EMG signal classification, they apply a 3-layers ANN with 7 neurons as input layers, 10 tan-sigmoid function neurons as hidden layer and 4 linear function neurons as output layer (Figure 5.). To prevent the ANN from overfitting, they divide the training data into three parts i.e. the training data (70%), the validation data (15%) and the testing data (15%). Early stopping method is also used to improve the generalization of the network. The experimental result shows that the average success classification rate of their system is 88.4% and 89.2% at the best time.

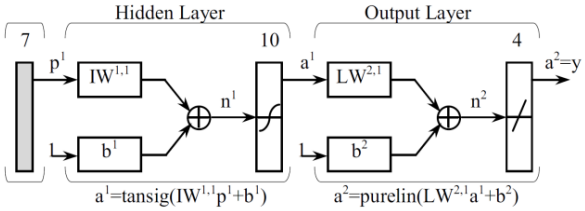

Figure 5. Architecture of Artificial Neural Network [6]

Marco E. et al. propose a real-time HGR approach based on the surface electromyography signals measured on the muscles of the forearm by the MYO armband [7]. The MYO armband (Figure 6.) is a commercial sensor that embodied with 8 dry and bipolar surface EMG pods. This sensor can measure the electrical activity of the muscles of the forearm, close to the elbow. It can also acquire the acceleration, angular velocity and orientation information of user's arm. The structure of their model consists five important modules i.e. the EMG signal collecting module, the preprocessing module, the feature extraction module, the classification module and the postprocessing module. Transmitted through Bluetooth, they obtain the EMG signal which is represented in N by 8 matrices with the 8 EMG pods of the MYO armband in the first module.

In the preprocessing step, they rectify the EMG signal measured in the first module by passing the signal through a low-pass Butterworth filter to obtain the envelop of the EMG signal, which contains the information of the movement of use's hand. After several steps of data processing, they define two kinds of feature matrices, one for training and one for testing, based on the data acquired form the preprocessing module in the feature extraction module. The k-nearest neighbor and the dynamic time warping (DTW) algorithm are used for classifying the EMG signal. To eliminated consecutive of the same label, they apply a time delay of one step in the postprocessing module. Finally, their gesture recognition system can distinguish five different hand gestures, including “fist”, “wave in”, “wave out”, “open”, and “pinch”, with a precision of 89.5%. But the experiment result shows that their module is highly user-specific. The performance of the system is less accurate in the general module with only 53.7% classification rate.

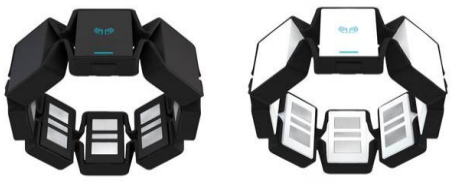

Figure 6. MYO armband

B. Vision Based Approach

1) 2-D Image Processing

Mandeep Kaur Ahuja and Amardeep Singh propose a gesture recognition scheme by using a database-driven HGR based upon skin color model approach and thresholding approach along with an effective template matching using PCA [8]. Like the other vision-based approaches (Figure 7.), their gesture recognition system consists of three parts i.e. the image acquisition part, the hand segmentation part and the feature extraction & gesture recognition part.

The images of gestures are acquired using a 8 megapixel real-aperture camera in constant background in the first part. Then they apply skin colour model in YCbCr colour space along with Otsu thresholding to separate hand gesture information from background. In the final step, principle component analysis (PCA) is used for feature extraction and matching. Their system can recognize four static hand gestures with 91.25% average accuracy and 0.098251 seconds average recognition time.

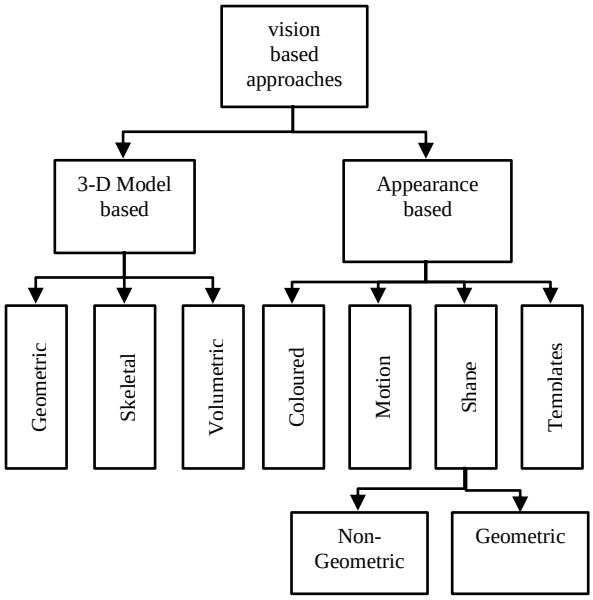

Figure 7. Categories of Vision Based Categories [8]

Hussain et al. develop an HGR system based on deep learning [9]. Their system can recognize six static hand gestures and eight dynamic hand gestures. The main steps of their HGR process include hand shape recognition, tracing of detected for dynamic gesture recognizing and transforming the data into the corresponding command.

For hand shape recognition, they train a CNN based classifier through the process of transfer learning over a VGG16 neural net which works as the pretrained model. In order to distinguish dynamic hand gesture, a hand tracing phase is applied in the second steps. In hand tracing phase, hand area is segmented out using HSV (Hue, Saturation, Value) skin color algorithm in a frame, followed by cropping blob area, the center of which is detected and traced. By comparing the coordinate of the centroid in several frames, the direction of dynamic gestures can be acquired thus the command of the dynamic gesture can be determined. The experiment result shows that the gesture recognition system can achieve an accuracy of 93.09%. The workflow of their system is presented in Figure 8. .

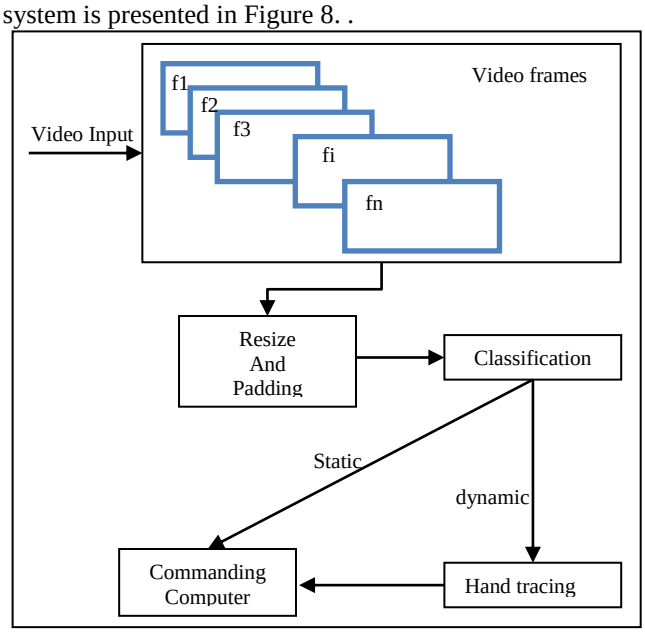

Figure 8. Workflow [9]

2) Depth sensor

Lu et al. propose a dynamic hand recognition scheme with Leap Motion Controller (LMC) [10] The framework of their system consists of two main parts: feature extraction and classification with the Hidden Conditional Neural Field (HCNF) classifier (Figure 9.). The features used for further classification is determined based on palm direction, palm normal, fingertips positions and palm center position, which can be acquired easily via LMC. For better classification, they propose two different features: the single-finger feature to deal with the mislabeling problem and the double-finger feature to distinguish the different types of interactions between adjacent fingertips.

Using LMC's specific API, they build two kinds of dynamic hand gesture datasets i.e. the LeapMotion-Gesture3D dataset which contains 12 gestures and the Handicraft-Gesture dataset, which consists of ten hand gestures originated from pottery skills. After the feature extraction procedure, a HCNF-based classifier is applied to recognize dynamic hand gestures, which can take two main factors into consideration for improving the accuracy: different kinds of features and complex underlying structure of dynamic hand gesture sequences. The experimental results show that their method achieved 95.0% recognition accuracy for the Handicraft-Gesture dataset and 89.5% for the LeapMotion-Gesture3D dataset.

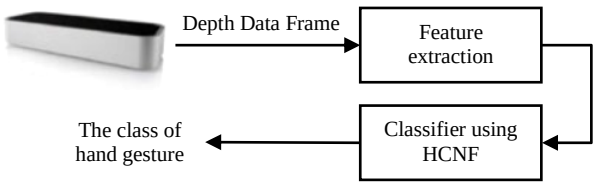

Figure 9. The systematic framework [10]

Nguyen-Duc-Thanh et al. develop a hand gesture system for Human-robot Interaction (HRI) based on a 2 - stages Hidden Markov Model [11]. The scheme of the approach is presented in Figure 10. . The human skeleton is detected from the sequence of image frame acquired via Kinect sensor for features extraction.

The skeletal feature vector is defined by relative difference between the current position and the starting position of the hands' and elbows' skeleton points. Then the feature vectors are feed to the first HMM for prime gesture recognizing. Based on the sequence of prime gestures recognized from the first HMM, the second HMM plays a role in recognizing the whole task. They implemented the whole model on an iRobot Create. The experiment result showed that in the gesture recognition part, the model can achieve an average accuracy of 95.33%.

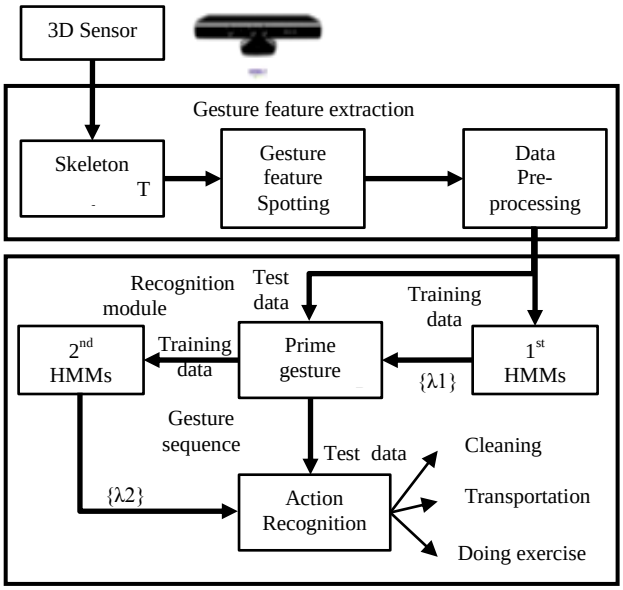

Figure 10. Proposed framework [11]

C. Comparison

Sensor based gesture recognition technique uses sensors to measure the position information of user's hands to recognize hand gesture. Acceleration sensor, flex sensor, surface acoustic wave sensor and infrared range sensor are also commonly used for measuring hands' kinematic information and other useful information. For the purpose of collecting raw hand gesture data, the user needs to wear an additional device such as data glove and arm band which are integrated with various kinds of sensors.

This may be the major drawback of sensor based gesture recognition method because of the inconvenience of connecting to the computer physically. Based on different types of the collected raw data, other problems like unable to extract appropriate features and cumbersome noise filtering may be occurred. However, compared to vision based technique, sensor based technique has little environment influence and high real time response. It can avoid the visual occlusion effectively at same time shorten the time costs for running complex image analyzing algorithm.

In comparison with sensor based gesture approach, on the one hand, vision based approach does not require users to wear any device thus leave more freedom for the users. On the other hand, vision based HGR method depend highly on the external environment and the corresponding algorithm are relatively complex. For 2D computer vision based gesture recognition approach, if the light condition is less harmonic or the skin color is close to that of the user's cloth, the accuracy of recognition will be affected badly.

With the emergence of depth image-based sensor, a new approach in HGR is get more and more attention. Pixels in a depth image carry information about the object's distance regard to the sensor. By considering the depth information, the image segmentation process is much convenient and immune to illumination changes which is a major bottleneck for 2D image segmentation methods. In recent years, many studies were conducted by using commercial depth sensors, such as Leap Motion Controller (LMC) or Microsoft Kinect sensor. The experiment results of their works show a promising feature of gesture recognition using depth sensors. But several challenges still exist, such as the online recognition of 3D gestures and distinguishing similar hand gestures with different meanings [12]. The precision of the depth sensor itself is a strong limitation of recognition accuracy too.

III. Application of Gesture Recognition in HRI

With the advancement of industrial automation, human robot interaction (HRI) has become an emerging field in recent years. Hand gestures can be effectively used to give commands to the robot which in tum can be employed in large number of applications. The following section focus on the implementation of gesture recognition in human-robot interaction and the structures of developed HRI system. Some related works will be discussed in detail.

A. Related Works

Ganapathyraju et al. develop an HRI system based on gesture recognition [13]. Their human-robot interaction system can be divided in to two sections, the hand image acquiring and processing section as well as the robot control section. In the first section, they use a webcam to obtain hand images and then process these images through several stages (Figure 11.) so that the message within hand gestures can be distinguished correctly. After decoding the meaning of user's gesture command by using image processing, they send the command to an ABB IRB 120 6 axis robot using serial communication, thus complete the mission of section two.

Their work emphasizes on the first part, the image processing part, in which they use the skin detection algorithm in YCbCr colour space to isolate the hand gesture form the rest of the environment for the purpose of hand contour detection. After processing the acquired images through skin detection, they apply morphological

technique (erosion and dilation) to filter the images and then acquire the contour of hand image by conducting contour detection algorithm. The final and the most important step of part one is using convexity hull and convexity hull defects to get the count of the fingers using the following algorithm (1)

$$
\begin{aligned}
&Count = \{1\}\\
&\text{If}(S_Y < B_Y \text{ or } D_Y < B_Y) \text{ and } (S_Y < D_Y) \text{ and}\\
&\quad (L_D > \text{box height}/n)\\
&\quad \text{Where, } L_D = \sqrt{(S_X - D_X)^2 + (S_Y - D_Y)^2}\\
&\quad Else\ Count = \{0\}\\
&Number\ of\ Fingers\ counted = \sum Count
\end{aligned}
\tag{1}
$$

where S_X is the start point, B_Y is the center point of the box, D_Y is the depth point, L_D is the length of the defect.

After converting gesture images into specific moment command for ABB robot, the remained question is to transmit that information to robot controller which is accomplished via serial cable connected the IRC5 ABB robot controller. Their method of gesture recognition through 2-dimensional image processing although effective, yet the accuracy of the system is affected by a number of factors which include background noise from background images and the angle of the tilt of the hand and its orientation.

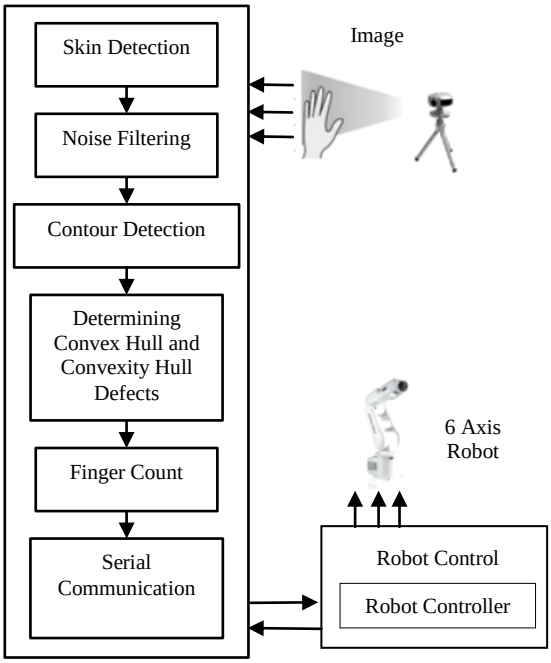

Figure 11. Proposed System Architecture [13]

Gu et al. build a human-robot communication system through a 3-D sensor Kinect produced by Microsoft corporation. Their gesture recognition system consists of three layers: hand detection, gesture Tracking and gesture recognition [14]. The detection layer and tracking layer are implemented with Kinect divers and the middleware called NITE [15]. In the first two layers, they use a Kinect sensor to capture RGB images as well as depth information of user's hand. Then the skeleton model of the user with fifteen joints can be detected by means of OpenNI driver and NITE. The angles of user's skeleton joint regard to his torso is used as features for detecting gesture information.

In the last layer, they applied Hidden Markov Models for gesture classification. Through segmentation and symbolization, they feed the angle information into each HMM of five distinct gestures to adjust the model parameters $\lambda = (A, B, \pi)$ to maximize $P(O|\lambda)$, where A is the state transition probability distribution, B is the observation symbol probability distribution, π is the initial state distribution and $P(O|\lambda)$ is the probability of observation sequence of the given model. Viterbi algorithm [16] is adopted for computing $P(O|\lambda)$ in recognition phase which is implemented on Robot Operate system (ROS) framework (Figure 12.). Their HRI system is applied on a mobile robot i.e. a Pioneer robot [17]. By using five gestures predefined for the system, they can control the mobile robot with an average accuracy around 85% for the person who provide the training data and 73% for the other subjects.

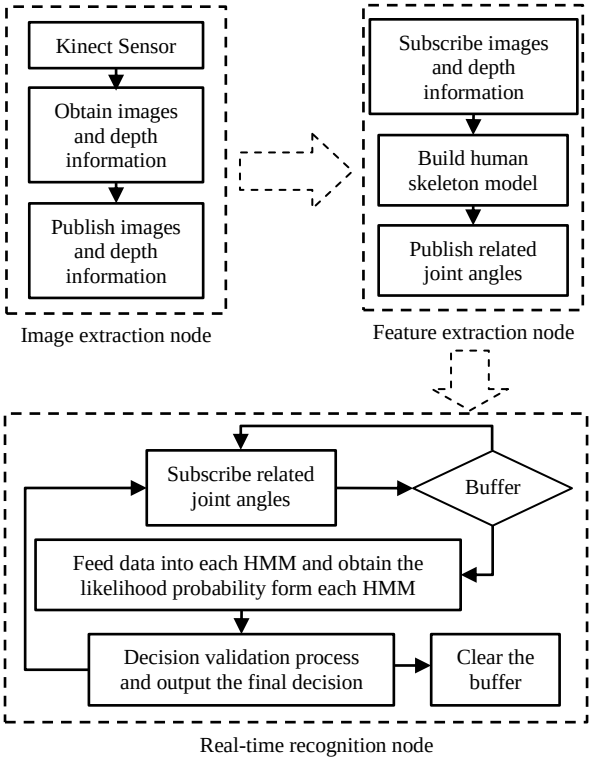

Figure 12. Flowchart of the recognition phase [15].

Ariyanto et al. present a project by using hand gesture to control a five degree of freedom robotic arm, SCORBOT [18]. Their HRI system consists of three main modules: vision processing, gesture recognition and robotic arm driver. In vision system, they use a camera mounted at the top of the shelf looking down on a black carpet to capture user's hand motion frame by frame using open source visual processing library (OpenCV). For each video frame, the row image is converted into a colour probability distribution image, on which they apply the Cam-shift Algorithm to track user's hand and

determine the region of interest (ROI). Then, only the pixels within the ROI will be considered in future process.

The second module of their project is split into two main parts: The Principal Component Analysis and the Artificial Neural Networks. In the first part, they build the eigen objects and then calculated the decomposition coefficients of the images collected in the first module. In the second part, they use a three layers feed forward neural networks and back propagation error training algorithm to classify the decomposition of coefficients vector from PCA.

The structure of the neural consists of 10 input neurons, 16 hidden neurons with tangent sigmoid activation function and 6 output units with linear activation function. Each output units corresponded to a static hand gesture. With a learning rate of 1.0, a momentum rate of 0.9 and a epoch parameter of 1000, using gradient descend back propagation, their artificial neural networks can achieve an recognition accuracy of 90-95% percent on the best time. In the last module, they design two different control type to control the robot arm i.e. real-time movements control mode and pre-defined movements control mode through which SCORBOT can do some simple job in real-time and can finish some simple pick-and-place work by pre-defined movements.

Luis-rodolfo et al. develop a flexible trajectory generation approach based on Microsoft Kinect sensor to generate new trajectories for an ABB IRB1600ID welding robot [19]. This trajectory generation system consists of two other main components i.e. a personal computer for data processing and an ABB IRC5 robot controller (Figure 13.). Their works mainly focused on data processing stage on PC which contains two major programs: a RAPID program called controller program for robot movement control and acquiring trajectory information from the second program named Kinect program (Figure 14.).

The Kinect program is written in C# language using Kinect for Windows SDK to interact with the Kinect sensor. The SDK provides a number of advanced image-processing functions, especially contains a skeleton extraction function, which plays an indispensable role in trajectory generation. In the program, they acquired user's skeleton model and extracted the 3D Cartesian coordinates of each hand in meters. By comparing the heights of the hands, different commands are sent to the controller program according to a sequence of pre-programmed actions.

If the height of the left hand is higher than that of the right the first time, the trajectory teaching procedure is initialized, and the welding robot returned to its home position. At the second time, the Kinect program suggests the controller program a new target with X, Y, Z coordinates of the user's right hand in meters and the robot responds with a bow. Another target point is obtained at the third time and at the fourth time, the robot will start welding along the trajectory determined by the last two target points.

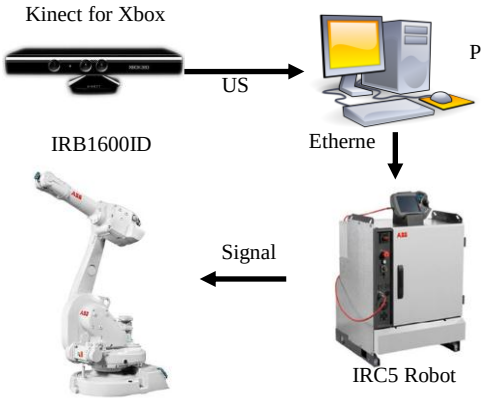

Figure 13. Main components [19].

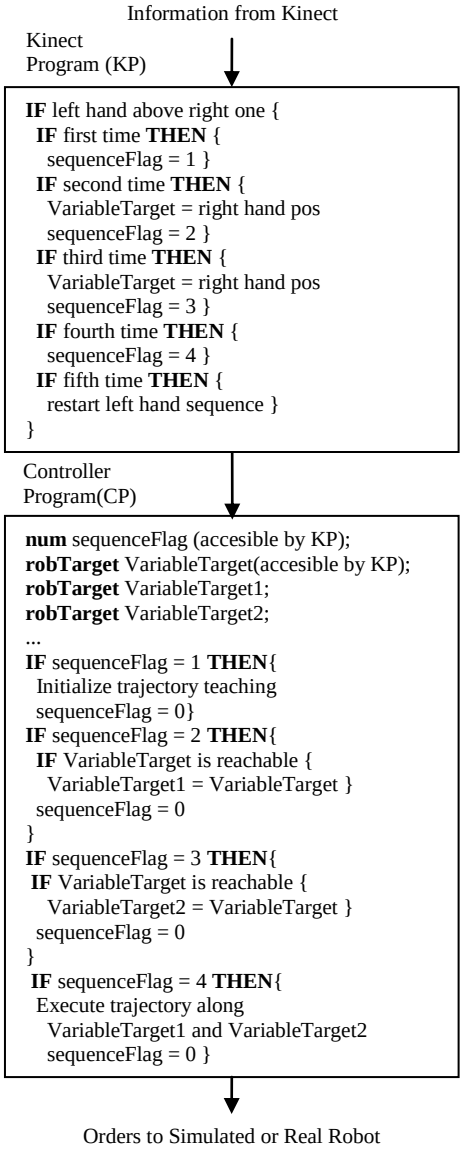

Figure 14. Programs interaction for teaching robot trajectories using the Kinect sensor [20].

B. Summary

Although the prospect of the gesture recognition based HRI is promising, current researches point out that there are still a lot of issues needed to be worked on before the

occurrence of a practical HRI model for industrial application. Issues like the real-time property of the gesture recognition system, the recognizing accuracy in an actual industry environment, the rationality and the practicality of the designed hand gesture set etc. should be considered in designing the gesture based HRI system. This indicates that the gesture recognition system based on depth sensor will get more and more attention because of its simplicity in extracting hand gesture information and the ability to adapt different environment without being affected by the lighting condition.

The demonstrated works successfully integrate gesture recognition with HRI. However, in the process of designing gesture commands, they do not take ergonomic factors into consideration. The proposed gesture set doesn't meet the user's demands in the real work situation. These gestures are either cumbersome to perform or incompatible with common body language, which in turn increase the cognitive workload and training costs significantly. Thus, hand gesture commands designed specifically for industrial purpose plays an indispensable part in the gesture recognition based HRI system.

IV. Methodologies of Gesture Recognition

From the previous works, we can make a conclusion that the scheme of gesture recognition generally contains three indispensable modules i.e. the gesture data collection module, the feature extraction module as well as the gesture classification module which classifies different gestures via appropriate classifier. Based on different approaches of gathering hand gesture information and different types and structures of the raw data, researchers develop distinct feature extraction methods, such as distance and angle from the endpoint of hand [20], orientation histogram [20], for further recognizing step. These features can be summarized into several categories which include spatial domain features, transform domain features, curve fitting-based features, histogram-based descriptors, and interest point-based descriptors [20].

In the hand gesture classification module, several classifiers are commonly used for recognizing the incoming gesture features and grouping them into predefined classes or by their similarity. We will focus on some frequently used pattern recognition algorithms applicable to HGR in this section.

A. K-means:

Means algorithm (Figure 15.) is the most simple and intuitive cluster analysis algorithm. Its purpose is to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, which is defined as the sum of the distances of all data points to their respective cluster centers.

The algorithm generally contains two steps: the assignment step and the update step. Given an initial class number of k, the algorithm selects k center points randomly. Then, each observation is assigned to the nearest cluster center according to their distance from the center point in the first step. In the second step, the center of each cluster is recalculated based on the average location value and then update. By alternating between these two steps, the algorithm has converged when the assignments no longer change. Generally, the process will stop under the condition of a user specified maximum number of iterations or within a distance threshold of the movement of cluster centers.

Cheng et al. design an automatic museum robotic guide which integrates image and gesture recognition technologies. In the gesture recognition module, they use k-means algorithm as the first phase of their hybrid machine learning-based framework to train historical data and filter outlier samples for preventing future interference in the recognition phase [21].

Input: k (the number of cluster),
D (a set of lift ratios)
Output: a set of k clusters
Method:
Arbitrarily choose k object from D as the initial cluster centers;
Repeat:
1. (re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;
2. Update the cluster means calculating the mean value of the objects for each cluster
Until no change;

Figure 15. The pseudo code for K-means clustering algorithm [22]

B. Dynamic Time Warping (DTW):

DTW is one of the most used measure of the similarity between two time series, and computes the optimal global alignment between two time series, exploiting temporal distortions between them. It was originally proposed in 1978 by Sakoe [23] and Chiba for speech recognition, and it has been used up to today for time series analysis such as temporal sequences of video, audio, and graphics data. Indeed, any data that can be turned into a linear sequence can be analyzed with DTW. Thus, DTW algorithm is widely used in matching the incoming gesture features with the predefined templates [24][25][26].

Given two time series v1=(a1,...,an), v2=(b1,...,bn), the algorithm defines a similarity matrix and fill the matrix with the differences of the two time series. Then, the greedy algorithm is applied to find the path between the bottom-left corner and the top-right corner with the minimum accumulated distance. The pseudo code of DTW algorithm is presented in Figure 16. .

C. Hidden Markov Model (HMM):

Hidden Markov Model (Figure 17.) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved states. Given a sequence of units (words, letters, morphemes, sentences, whatever), HMM can compute a probability distribution over possible sequences of labels and choose the best label sequence [27]. The algorithm is always used in recognizing dynamic hand gestures.

```
DTW(v1, v2) {
//where the vectors v1 =(a1 , ... , an), v2 =(b1 , ... , bm) are the time series
with n and m time points
    Let a two-dimensional data matrix S be the store of similarity
measures
such that S[0 , ... , n , 0 , ... , m], and i , j , are loop index , cost is an
integer.
    //  initialize the data matrix
    S[0 , 0] := 0
    FOR i := 1 to m DO LOOP
        S[0 , i] :=
    END
    FOR i := 1 to n DO LOOP
        S[i , 0] :=
    END
     // Using pairwise method, incrementally fill in the similarity
matrix
     with the differences of the two time series
    FOR i := 1 to n DO LOOP
        FOR j := 1 to m DO LOOP
        // function to measure the distance between the two points
            cost := d(v1[i] ,v2[j])
            S[i , j] := cost + MIN(S[i - 1 , j]),      // increment
            S[i , j - 1]    // decrement
            S[i - 1 , j -1]    // match
        END
    END
    Return S[n , m]
}
```

Figure 16. Pseudo code of dynamic time wrap algorithm [27].

The modelling of a gesture sequence using HMM involves two steps: feature extraction and HMM training. In the feature extraction step, a particular dynamic gesture can be represented by a number of feature vectors. Each feature vector describes the state of the hand corresponding to a specific state of the gesture. Depending on the complexity of the gesture, the number of such states can be various. In the second step, the feature vectors are fed in to HMM for gesture classification.

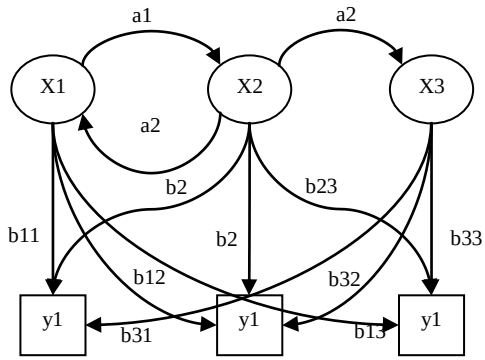

Figure 17. Probabilistic parameters of a hidden Markov model (example), X - states, y - possible observations, a - state transition probabilities, b -output probabilities

Xu et al. propose a new algorithm based on HMM and DTW to enhance efficiency and accuracy of dynamic HGR [28]. Dai et al. propose a novel method for simultaneous gesture segmentation and recognition based in HMM with an average segmentation accuracy of 93.88% and an average corresponding recognition accuracy of 92.22% [29].

D. Support Vector Machine (SVM):

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In two-dimensional space this hyperplane is a line (Figure 18.) dividing a plane in two parts where in each class lay in either side. This mapping from lower to higher dimensional spaces makes the classification of the input data simpler and more accurate.

Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other. Therefore, SVM was originally designed for two-class classification. For extending it to multi-class classification, one approach is to create several one-vs-all classifiers. Another way is to formulate SVM as a K-class classification problem, based on the classical paper by Weston and Watkins [30]. They compare it with K-binary or one-vs-all method show that it results in less support vectors. The current multi-class classification methods are mainly the following two situations, i.e. ‘one-against-all’ and ‘one-against-one’ methods. After this extending, SVM can be applied on multi-class classification system like HGR.

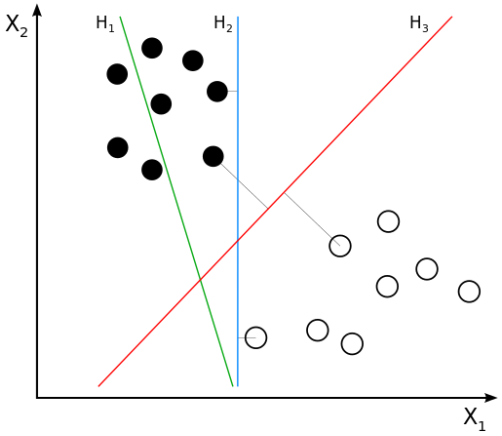

Figure 18. H1 does not separate the classes. H2 does, but only with a small margin. H3 separates them with the maximum margin.

Oh et al. propose a method to recognize the human gesture using binary decision tree and Multi-class Support Vector Machine. Feng et al. develop a static HGR model based on Based on HOG Characters and Support Vector Machines [31].

E. Artificial Neural Network (ANN)

Artificial neural networks are one of the main tools used in machine learning. As the “neural” part of their name suggests, they are brain-inspired systems which are intended to replicate the way that we humans learn. Neural networks consist of input and output layers, as well as (in most cases) a hidden layer consisting of units that transform the input into something that the output layer can use (Figure 19.). They are excellent tools for finding patterns which are far too complex or numerous

for a human programmer to extract and teach the machine to recognize.

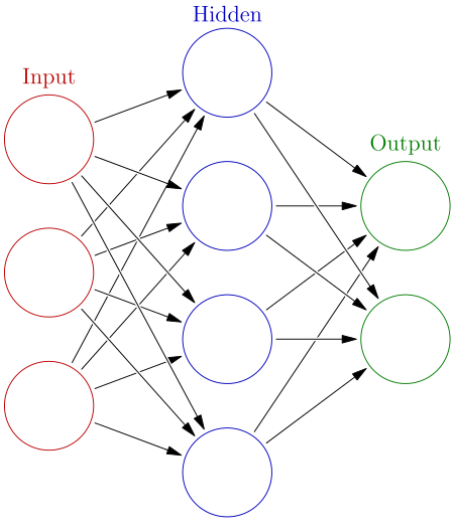

Figure 19. An artificial neural network is an interconnected group of nodes, akin to the vast network of neurons in a brain.

There are multiple types of neural network. Each of them comes with their own specific use cases and levels of complexity, such as feedforward neural network, recurrent neural network, convolutional neural networks, Boltzmann machine networks, Hopfield networks, and a variety of others. Artificial neural network can be used as a supervised classifier for gesture recognition. After training the ANN with a set of labeled input patterns, the ANN classifies new input patterns within the labeled class. It can be used to recognize static [6][32] and continuous hand gestures [33][34].

V. Design of Gesture Set of HRI

As we mentioned earlier, in the process of designing gesture commands, many factors related to the design of the gesture recognition system and the practicality of the gesture set must be taken into consideration. Hand gesture commands designed specifically for industrial purpose plays an indispensable role in the gesture recognition system.

There are some factors in industrial scenario that should be considered in designing hand gesture set, such as the simplicity of performing the gesture, the uniqueness or difference of distinct gestures and so on. Researches of designing hand gesture set for industrial purpose have been performed by several scholars. In this section, we will make a summary of current related works and discuss the principles of designing industrial hand gesture commands.

A. Related Work

Barattini et al. design a gesture set which contain 12 basic gestures and 7 supplementary gestures [2]. They suggest several principles in designing gestures for industry. First, the gestures must be easy to make, as close as possible to the common use of finger/hand/arm gestures (i.e. "nature"). Second, the gestures must be clearly distinguishable one from another. Third, the gestures must easy to remember to minimize training time. Fourth, the gestures must be different from movements done to perform work tasks or gesticulation done while talking.

Flowing these principles, most of their designed gestures can be performed dynamically with only one arm while other gestures are combinations of movement of arm, hand and fingers. Of all 12 gestures, two of them are special, according to Barattini, the "Identification" gesture and the "Change" gesture (Figure 20.) which solve the problems like manager identification and work-shifting. They make a evaluation on the capacity of a computer to distinguish between the gestures using Microsoft Kinect SDK for feature extraction and Dynamic Time Warping for gesture recognition. The experiment result shows their gesture set has a low confusion.

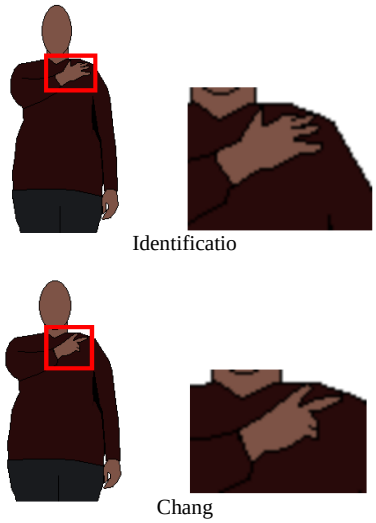

Figure 20. The "Identification" gesture and the: "Change" gesture described in the text [2].

Gleeson et al. present a gestural communication lexicon for human-robot collaboration in industrial assembly tasks based on observations of the communication between human pairs [35]. They aim at propose a gesture set that can both be used on human and robot partner. They divide the actions required for industrial assembly tasks into three classes: part acquisition, part manipulation, and part operations. Then, they design and conduct a human observation experiment to determine what terms need to be communicated and the gestures most naturally used to communicate these terms. The experiment result shows several interesting facts.

First, the gestures (see Figure 21.) used during the task is rather physically simple. Nearly all gestures can be executed with one hand, which makes the gesture set suitable for implementation on an arm robotic assistant.

Second, a simple gesture set can convey diverse meanings. Based on different work scenarios, a gesture communicates different ideas, which lessens the burden on designing and implementing the gestures and simplify the task of gesture recognition.

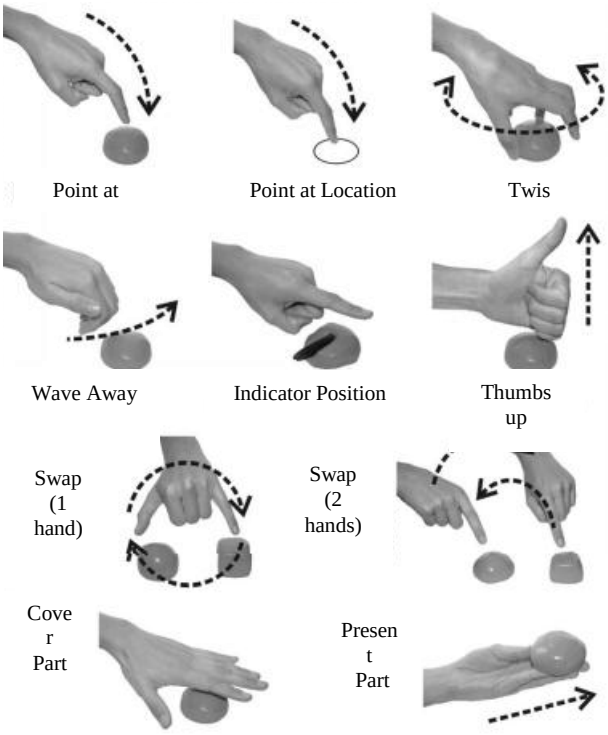

Figure 21. The proposed gesture lexicon [36]

In the third place, they find two basic "sentence structures" for gestural communication, communications that specified both an object and an action, and those which specified only an object and relied on an implied understanding of the desired action, which require the robotic assistant to estimate the intent of the human operator based on ambiguous communication plus the state of the task.

Fourth, the interpretation of gestures is highly context sensitive, which provides a challenge for HRI. At last, the gestures are highly intuitive and broadly applicable. However, the context-sensitivity of gestural communication does limit the applicability of their lexicon.

B. Discussion

The above works give us some feasible suggestions when designing a gesture set. We can sum up these suggestions and define several generic principles for designing gestures for industrial purpose.

The first and the most important principle is that gestures used in HRI must be natural to execute and easy to remember. Second, different gestures must be clearly distinguishable from one another, and different meanings of a same gesture must also be distinguishable according to the specific work scenario. Third, the performed gesture must be different from the movement of the hand when talking or working which can prevent HRI from confusion. Fourth, gestures that frequently used should be designed as single-hand gestures for convenience. Finally, we must consider the workload in developing gesture recognition system when design gesture set, which means, instead of planning a complicated generic gesture set, we can design several simple subsets for different work scenario, and gestures in different subset can be identical, yet with distinct meaning. Thus, when the work changes, the only thing user need to do is to switch the gesture subset.

VI. Conclusion and Future Works

HGR (HGR) is a natural way of Human Machine Interaction and has been applied on different areas. In this paper, we discussed works done in the area of applications of hand gestures recognition in industrial robots where focus is on the processing steps and techniques in recognizing hand gestures. We elaborated several related works in the area of HGR based on different approaches including sensor based approach and vision approach. What we noticed in comparing these two approaches is that depth sensor based HGR method is convenient and powerful. It can both be applied on dynamic HGR and static HGR. Then concerning works of implementation of HGR in industrial scenario were discussed and analyzed. Pattern recognition algorithms that effectively used in HGR like k-means, DTW etc. were briefly introduced as well. Finally, after reviewing some related works, we proposed several principles of designing industrial hand gesture commands. In the future we will work in the area of individual gesture designing and its application on industry robot, as work done in this area are rare. We will develop an HGR system based on a combination of depth image approach and 2D image processing, and implement the HGR system on an industry robot.

Acknowledgement

This research is supported by the Guangdong Innovative and Entrepreneurial Research Team Program (Grant No.2014ZT05G132), Shenzhen Peacock Plan (Grant No.KQTD2015033117354154), the Major Projects of Guangzhou City of China (Grant No.201704030091,201607010041), the Major Projects of Guangdong Province of China (Grant No.2015B010919002), the Major Projects of Dongguan City of China (Grant No.2017215102008), the Nansha District International Science and Technology Cooperation Project of Guangzhou City of China (Grant No.2016GJ004).

References

[1] A. Kendon, *Gesture: Visible Action as Utterance*, UK: Cambridge University Press, 2004.

[2] P. Barattini, C. Morand, and N. M. Robertson, "A proposed gesture set for the control of industrial collaborative robots," in *Proc. 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, Paris, 2012, pp. 132-137.

[3] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 3, pp. 311-324, May 2007.

[4] N. Lv, X. Yang, Y. Jiang, and T. Xu, "Sparse decomposition for data glove gesture recognition," in *Proc. 2017 10th International Congress on Image and Signal Processing, BioMedical*

Engineering and Informatics (CISP-BMEI), Shanghai, 2017, pp. 1-5.
[5] Y. Ge, B. Li, W. Yan, and Y. Zhao, "A real-time gesture prediction system using neural networks and multimodal fusion based on data glove," 2*018 Tenth International Conference on Advanced Computational Intelligence (ICACI)*, Xiamen, 2018, pp. 625-630.
[6] M. R. Ahsan, M. I. Ibrahimy, and O. O. Khalifa, "Electromygraphy (EMG) signal based HGR using artificial neural network (ANN)," in *Proc. 2011 4th International Conference on Mechatronics (ICOM)*, Kuala Lumpur, 2011, pp. 1-6.
[7] M. E. Benalcázar et al., "Real-time HGR using the Myo armband and muscle activity detection," in *Proc. 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)*, Salinas, 2017, pp. 1-6.
[8] M. K. Ahuja and A. Singh, "Static vision based HGR using principal component analysis," in *Proc. 2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education (MITE)*, Amritsar, 2015, pp. 402-406.
[9] S. Hussain, R. Saxena, X. Han, J. A. Khan and H. Shin, "HGR using deep learning," in *Proc. 2017 International SoC Design Conference (ISOCC)*, Seoul, 2017, pp. 48-49.
[10] W. Lu, Z. Tong and J. Chu, "Dynamic HGR With Leap Motion Controller," *IEEE Signal Processing Letters*, vol. 23, no. 9, pp. 1188-1192, Sept. 2016.
[11] N. Nguyen-Duc-Thanh, S. Lee and D. Kim, "Two-Stage Hidden Markov Model in Gesture Recognition for Human Robot Interaction," *International Journal of Advanced Robotic Systems*, vol. 9, no. 2, p. 39, 2012.
[12] H. Cheng, L. Yang and Z. Liu, "Survey on 3D HGR," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 9, pp. 1659-1673, Sept. 2016.
[13] S. Ganapathyraju, "HGR using convexity hull defects to control an industrial robot," in *Proc. 2013 3rd International Conference on Instrumentation Control and Automation (ICA)*, Ungasan, 2013, pp. 63-67.
[14] P.Kathuria, A.Yoshitaka, "HGR by using logical heuristics," *Japan Adv.Inst. of Science and Tech. School of Info.Science, Report*, 2012.
[15] Y. Gu, H. Do, Y. Ou, and W. Sheng, "Human gesture recognition through a Kinect sensor," *2012 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Guangzhou, 2012, pp. 1379-1384.
[16] *Openn.* [Online]. Available: http://www.openni.org/Documentation/
[17] M. S. Ryan and G. R. Nudd, "The viterbi algorithm," Coventry, UK, Tech. Rep., 1993.
[18] *Adept Mobilerobots.* [Online]. Available: http://www.mobilerobots.com/ResearchRobots/PioneerP3DX.aspx
[19] G. Ariyanto, P. K Patrick, H. W. Kwok, G Yan, "HGR using neural networks for robotic arm control," in *Proc. the National Conf. On Computer Sci & Information Technology*, Faculty of Computer Science, University of Indonesia, Jan 29-30, 2007.
[20] R. Mendoza-Garcia, L. Landa-Hurtado, F. Mamani-Macaya, H. Valenzuela-Coloma, and M. Fuentes-Maya, "Kinect-based trajectory teaching for industrial robots," in Pan-American Congress of Applied Mechanics, Santiago, Chile, 2014.
[21] J. S. Sonkusare, N. B. Chopade, R. Sor, and S. L. Tade, "A Review on HGR System," *2015 International Conference on Computing Communication Control and Automation*, Pune, 2015, pp. 790-794.
[22] T. Maung, "Real-time hand tracking and gesture recognition system using neural networks," PWASET, vol. 38, pp. 470-474, 2009.
[23] K. Chakraborty, D. Sarma, M. K. Bhuyan, and K. F. MacDorman, "Review of constraints on vision-based gesture recognition for human – computer interaction," in *IET Computer Vision*, vol. 12, no. 1, pp. 3-15, 2 2018.
[24] F. Cheng, Z. Wang, and J. Chen, "Integration of open source platform duckietown and gesture recognition as an interactive interface for the museum robotic guide," *2018 27th Wireless and Optical Communication Conference (WOCC)*, Hualien, 2018, pp. 1-5.
[25] P. Zhou and K. Chan, "A Model-based multivariate time series clustering algorithm," *Lecture Notes in Computer Science*, pp. 805-817, 2014.
[26] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoustic Speech and Signal Processing*, vol. 26, pp. 43-49, 1978.
[27] H. Choi, E. Kim, and T. Kim, "A DTW gesture recognition system based on gesture orientation histogram," *The 18th IEEE International Symposium on Consumer Electronics (ISCE 2014)*, JeJu Island, 2014, pp. 1-2.
[28] C. Hang, R. Zhang, Z. Chen, C. Li, and Z. Li, "Dynamic gesture recognition method based on improved DTW algorithm," *2017 International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII)*, Wuhan, 2017, pp. 71-74.
[29] S. Bodiroža, G. Doisy, and V. V. Hafner, "Position-invariant, real-time gesture recognition based on dynamic time warping," *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Tokyo, 2013, pp. 87-88.
[30] S. Fong, "Using hierarchical time series clustering algorithm and wavelet classifier for biometric voice classification," *Journal of Biomedicine and Biotechnology*, vol. 2012, pp. 1-12, 2012.
[31] J. Dan and J. H. Martin, "Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition," Prentice Hall series in artificial intelligence (2009): 1-1024.
[32] X. Zhang, J. Wang, X. Wang and X. Ma, "Improvement of Dynamic HGR Based on HMM Algorithm," *2016 International Conference on Information System and Artificial Intelligence (ISAI)*, Hong Kong, 2016, pp. 401-406.
[33] Y. Dai, Z. Zhou, X. Chen, and Y. Yang, "A novel method for simultaneous gesture segmentation and recognition based on HMM," *2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, Xiamen, 2017, pp. 684-688.
[34] J. Weston, C. Watkins, "Multi-class support vector machines," in *Proc. European Symp. Artificial Neural Networks (ESANN)*, 1999, pp. 219 – 224
[35] K. Feng and F. Yuan, "Static HGR based on HOG characters and support vector machines," *2013 2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation (IMSNA)*, Toronto, ON, 2013, pp. 936-938.
[36] B. Gleeson, K. MacLean, A. Haddadi, E. Croft, and J. Alcazar, "Gestures for industry Intuitive human-robot communication from human observation," in *Proc. 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Tokyo, 2013, pp. 349-356.

Qujiang Lei works as an Associate Professor at the Intelligent Robot and Equipment Center, Guangzhou Institute of Advanced Technology (GIAT), Chinese Academy of Sciences, Guangzhou, China. He received the Doctoral Degree in 2018 from Delft University of Technology, Delft, The Netherlands. His research interests include intelligent robots, collaborative robots, robotic systems integration, robot vision, artificial intelligence, machine learning and human-computer interaction.

Hongda Zhang is a graduate student in the school of mechanical engineering and automation at Harbin Institute of Technology. His research interests include intelligent robots, robot vision and human-robot interaction.

Yang Yang is a graduate student in the college of Big Data and Information Engineering at Guizhou University. Her research interests include computer vision and image processing.

Yue He received the B.S. degree in Communication Engineering from University of Northeast Electric Power University, China, in 2013. She is currently working toward the M.S. degree with the Institute of Communication Engineering, Department of Big Data and Information Engineering, Guizhou University, China. Her latest research interests include robotics vision and machine learning.

Yang Bai is a graduate student and currently pursuing the M.E. degree in Intelligent Building at Xi'an University of Architecture and Technology, Xi'an, China. His research interests include motion recognition, human robot interaction, collision detection and obstacle avoidance.

Shoubin Liu works as an associate professor at the school of Mechanical Engineering and Automation at Harbin Institute of Technology, Shenzhen. He received his Doctoral Degree in 2000 from City University of Hong Kong. His research interests include Opto-mechatronics machine, Precision instrument and intelligent robots.