International Journal of Computer (IJC)
ISSN 2307-4523 (Print & Online)

http://ijcjournal.org/

Hand Gesture Detection and Recognition System: A Critical Review

Phyu Myo Thwe[a]*, May The` Yu[b]

[a,b]*University of Computer Studies, Mandalay, Myanmar*
[a]*Email: phyumyothwe@ucsm.edu.mm* , [b]*Email: maytheyu@ucsm.edu.mm*

Abstract

Hand gesture recognition is used enormously in the recent years for interact human and machine. There are many type of gestures such as arm, hand, face and many other but hand gestures give more meaningful information than other types of gestures. There are many techniques for hand gesture recognition, such as color marker approach, vision-based approach, glove-based approach and depth-based approach. The main purpose of gesture recognition system is to develop a useful system which can recognize human hand gestures and used them to control electronic devices. This paper reviewed the most common used hand gesture recognition methods, tools and analysis the strength and weakness of these methods, and lists the current challenging problems of hand gesture recognition system.

Keywords: Hand gesture; segmentation; feature extraction; recognition; Human-computer-interaction.

1. Introduction

Recently, hand gestures have become the most critical part to communicate between human and machine. For example, hand gesture recognition is recently used to replace for the mostly deployed human computer interactive devices such as joystick, keyboard, mouse, etc.

However, according to the complexity, diversity, ambiguity and uncertainty of hand gesture, hand gesture recognition has been becoming the most challenging research topic.

---

* Corresponding author.

The main challenges that need to solve as research issue include:

- Illumination condition is the most sensitivity for hand gesture recognition of vision based system.
- Complex backgrounds, dynamic backgrounds are also main difficult for hand gesture recognition.
- The different size of the user's hand is another issue.
- The multiple gestures in the same background and different viewpoints.
- The several persons contains in the sense other than the real subjects.

2. Literature Survey Of Current Research

Chen and his colleagues (2014) used finger segmentation and background subtraction method for real-time hand gesture recognition and the accuracy was achieved as 96.69%. If there is moving objects color similar to that of the skin then degrade the performance of the system. Nevertheless, the machine learning algorithms can be used to extract the hand from the background and Time of Flight cameras provide the depth information that can improve the process of hand detection. This work can get the best results to address the clutter background problem and to develop the robust of hand detection when use TOF cameras and machine learning as future work [1].

The authors (2016) studied dynamic time warping algorithm for both static and dynamic hand gesture recognition on depth information. The rate of accuracy was obtained 92.4% over 55 gestures. This paper implemented a novel algorithm to improve the scanning time to identify the first pixel on the hand contour locate the fingertips by using K-curvature algorithm and then directional search algorithm for identification of the entire hand contour. The very complex background, illumination conditions, skin color and clothing have little impact on the resulted the proposed solution. This work was encountering some problems during testing when one of the users wore a bracelet. In this paper, as future improvements, an additional module will be added to allow for a more accurate adaption of the size of a user hand instead of the double training at different depths that is currently implemented in the system. This would improve the performance at the level of recognition stage. In the same direction, a different version of DTW will be evaluated for future speeding up the gesture recognition [2].

Bao and his partners (2016) implemented for tiny hand gesture recognition system by using deep convolutional neural network without any localization and achieved accuracy up to 97.1% with simple background and 85.3% with the clutter backgrounds. This strategy has strength for the creation of the training dataset and the own learning procedure because bounding boxes specifying hand regions are not necessary. This approach is very appropriate for the practical and real-time recognition of gestures and it is more efficient than other approaches both in space and time. According to this paper, the deep learning approach is more appropriate in object detection and pattern recognition than other algorithm because CNN can extract low level features, middle level features and high level features [3].

Soe and her partner (2018) used faster region-based convolutional neural network for real time hand pose recognition. In this system used 10 types of postures to control VLC Media Player. The average accuracy is

89.95% on NUS Hand Pose Dataset and 86.12% on images from webcam. Nevertheless, the only drawback cannot detect very tiny hand objects. The system can also control other applications on personal computer as extend work [4].

The authors (2018) presented Hough transform and neural network for spatio-temporal approach. In the spatiotemporal feature extraction contains spatial features of the static hand gesture based on the skin color; geometrical features and Fourier descriptor of temporal features for the dynamic gesture by using Hough transform. The accuracy of this research is 94 % on Cambridge database and 98 % on Sebastien database [5].

Yingxin and his colleagues (2016) implemented robust hand gesture recognition system by using modified CNN. The data preprocessing also performed using canny edge detection which effective to remove the influence of illumination variance. According to the literature, some paper not include preprocess but preprocess can reduce the amount of data and remove the unwanted noise and region when preprocessing used before the feature extraction. The system achieved 98.2% of accuracy on self-constructed dataset. CNN and HMM can integrate to model dynamic hand gesture for robust spatio-temporal hand gesture recognition as future work [6].

The authors (2018) studied novel finger skin pixel algorithm that can quickly and easily recognize in the complex background and obtained accuracy as 98% [7]. Dehankar and his partners. (2017) presented real time end point identification method (RTEPI) and obtained accuracy 100% .But, on simple background tested only [8].

Dardas and his colleagues (2011) implemented RTGRS using bag-of-features and support vector machine techniques. In this system includes detecting and tracking hand in clutter background using skin detection and hand posture control comparison algorithm after face subtraction, recognizing hand gestures via. In the training stage, SIFT algorithm to extract features and then used vector quantization (VQ) technique which clusters the key points (features) using K-means clustering algorithm. This system recognized 10 postures with recognition accuracy 96.23%. The number of training image, the quality of webcam, and choosing number of clusters to build cluster model are very important factors for this system [9].

Chaikhumpha and his partners (2018) studied condensation and hidden markov models for real time two hand gesture recognition. In this paper, the processes are classified into two categories: hand tracking and gesture cognition to recognize hands. In the first process, condensation density propagation (condensation algorithm) is used to localize and track hands (centered at the center of palms) when they are moving. This system recognized 8 gestures used and the accuracy achieved 96.25% [10].

The authors (2016) implemented hand gesture recognition using two-level speed normalization feature selection and classifier fusion. In this paper, the author used two-level speed normalization procedure using DTW and Euclidean distance-based technique. ANOVA and IFS was used to choose the best features from the existing 44 features. The accuracy of 94.78% was achieved [11]. Hussain et.al. (2017) implemented hand gesture recognition using deep learning and with an accuracy of 93.09%. In this paper, the author was used VGG 16 pre trained model for recognize the hand [12].

3. Comparison Of Methods

Based on the literature review of several papers, comparison of most prominent papers is given below in Table I.

Table I: Comparison of major important research work

Research Article	Finding and Limitation
Plouffee and his colleagues (2016)	1. Implemented on depth data and achieved 92.4% of accuracy. 2. K-curvature algorithm and Dynamic Time Wrapping Algorithm was used. 3. The some problems were encounter when users were bracelet during testing. 4. An additional module can be added for more accurate adaptation of the size of a user hand instead of the double training at different depths as future work.
YingXin and his colleagues (2017)	1. Proposed to recognize the very small hand gesture without any segmentation and detection stage using deep learning. 2. Recognition accuracy was reported as 97.1 with simple background and 85.3% with clutter background. 3. Can't evident improvement in performance when applied to the original RGB image without any pre-processing. 4. Need more robustness and efficiency with complex and outdoor senses.
Soe and his colleagues (2018)	1. Faster Region-Based Convolutional Neural Network was used and achieved 86.12 % of real time webcam image. 2. Can't detect very small hand. 3. Can extend to control other application.
Patil and his colleagues (2018)	1. Implemented using Hough Transform and Neural Network and obtained accuracy 94% on Cambridge dataset and 98% on Sebstien Dataset. 2. This system is no dependent on wide range of variation in color, illumination, orientation. 3. The work can be improved computational efficacy and adapting for real time.
Yingxin and his colleagues (2016)	1. Not influenced by rotation, individual, scale, illumination and translation etc. 2. Edge RGB and CNN was used. 3. The result more accurate original RGB+CNN than edge RGB+CNN. 4. The clutter backgrounds were not tested.

4. Data Collection For Hand Gestures

There are basically three ways to collect the raw data for input of hand gesture recognition system.

- The first one is used color glove or data glove to collect the raw hand data called glove based approach. In data glove approach, the main drawbacks are expensive because sensor node, heavily, naturalness by using data glove. The weakness of glove user has to wear the glove every time. But, the color gloves are inexpensive; no sensors are embedded in or outside the gloves and robust method for hand gesture recognition system.
- The second way used one or more camera to collect raw input hand gesture called vision based approach. The strength of this approach is natural and more convenient for communication. But, easily affected by complex background.
- The last approach is hybrid approach to collect raw hand data by combining the above two methods.

5. The Segmentation Method Of Hand

Efficient the segmentation of hand is the main part of the human hand gesture recognition system because the segmentation of hand is mostly affected by various illumination conditions, static clutter background, dynamic background, and multiple gestures in the same background. In the segmentation stage, the data can be lost due to the inappropriate segmentation. So, the correct gesture cannot input to the gesture recognition system that will lead to incorrect recognition. The next steps mostly depend on the hand segmentation accuracy. The most common used hand segmentation techniques are as following.

A. Skin Color Segmentation
B. Background Subtraction
C. Contour Matching

A. Skin Color Segmentation

The segmentation with skin color is often used in hand gesture segmentation procedure. The researcher's mostly used color spaces are RGB, YCbCr and HSV. RGB color space is mixes three kinds of color information. So, skin range and luminance are highly correlated. However, YCbCr and HSV color space have better clustering features than color space of RGB. Luminance Y and color chrominance CbCr are obtained through linear transformation of RGB primary color. YCbCr color space is discrete space which gets better clustering results and computational efficiency [13]. The YCbCr color model is the most suitable for human skin color segmentation. HSV color space is continuous space. The main weakness of segmentation with skin color can be misclassified if the background contains objects color like the skin color.

B. Background Subtraction

The background subtraction method is also very simple frame differencing method [14]. In this method, the moving target can define as the difference between the current image and background image. The strength of

this method fast, easy to implement, effective detection and can also provide the complete information feature data of the target. The disadvantage of this method is particularly sensitive to the change in dynamic senses, such as indoor lighting gradually changes [15].

C. Contour Matching

The sequence of pixels that represent line/curve of the object in an image called contour [14]. This method is widely used for image objects classification and recognition. The contour matching is more suitable than template matching because contour matching algorithm is extracted to only the edge points rather than the whole image in template matching. The advantages of this method obtained sharp matches, faster, invariant under imaging transformations like rotation, intensity scaling and translation. The weakness of this method in recognition is tiny number of pixels is involved. So, the influence of each pixel may hinder recognition [16].

6. Hand Detection And Feature Extraction Model

In the feature extraction step, the most important features are identified from the segmented hand. After hand segmentation, the feature extraction is the second important part because the optimal set of features will obtained the best results of recognition performance.

According to [5] feature points are extracted based on skin color features, geometrical features and Fourier descriptor features as spatial features. After that, the temporal features are also extracted by using Hough transform. Finally, the spatial features and temporal features are combined as features vector.

7. Various Techniques For hand Gesture Recognition System

In the recognition phase, the features take as input and produce the output as class labeled which are required output gestures. The proper selection of gesture parameters of features can be affected the accuracy of the recognition. The most common used hand gesture recognition techniques are:

A. Dynamic Time Warping (DTW)
B. Artificial Neural Network (ANN)
C. Support Vector Machine (SVM)

A. Dynamic Time Warping

Dynamic Time Warping (DTW) deals with both static and dynamic aspects of gestures. DTW is used for a double role: to identify candidate gestures and to validate the similarity [2]. DTW algorithm used to compare the similarity between the capture gesture and gesture in the pre-trained data base. The advantage of this technique is time-compressing.

B.Artificial Neural Network

Many researchers already worked on gesture recognition by using neural network. as a classifier has been

utilized multilayer feed forward network that takes spatiotemporal feature vector as input. ANN is configured with one input layer, two hidden layer and output layers with 32-100-60-9 respectively [5]. The most widely used neural network algorithm is the feed forward back propagation algorithm in pattern recognition. The weakness of the artificial neural network is time delayed.

C. Support Vector Matching (SVM)

Support Vector Machine is also extremely popular algorithm which often used by many researchers as a classifiers. It is also supervised learning methods which can be used for both recognition and regression challenges.

It works really well with clear margin of separation. But, the weakness of this method is required higher training time when have large dataset and has more noise in the dataset. According to [9] bag-of-features is give as inputs vector for a multiclass SVM to classifier the types of gestures.

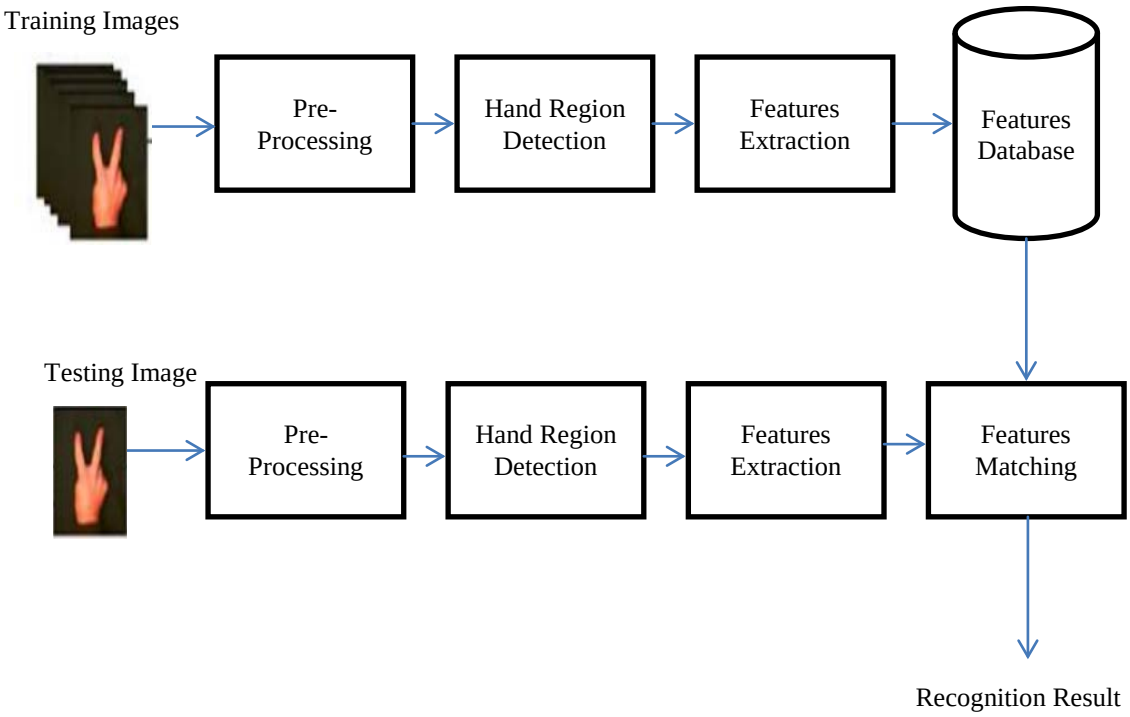

Figure 1: General Hand Gesture Recognition Stages

8. Conclusion

In this paper we have given a comprehensive survey of various hand gesture recognition systems. The principal component of the hand gesture recognition system are includes segmentation and tracking hand from the background and then feature extracted from the segmented hand image using various algorithms and finally, recognized the types of hand gestures by using recognition methods.

The most widely used hands gesture recognition techniques are review include ANN, DTW and SVM both of their strength and weaknesses are discussed.

9. Declaration

All author(s) have disclosed no conflicts of interest. The project was self-funded.

References

[1]. Z. H. Chen, J. T. Kim, J. Liang, J. Zhang, & Y. B. Yuan (2014). "Real-time hand gesture recognition using finger segmentation". The Scientific World Journal, 2014.

[2]. G. Plouffe, & A. M. Cretu (2016). "Static and dynamic hand gesture recognition in depth data using dynamic time warping". IEEE transactions on instrumentation and measurement, 65(2), 305-316.

[3]. P. Bao, A. I. Maqueda, C. R. del-Blanco, & N. García (2017). "Tiny hand gesture recognition without localization via a deep convolutional network". IEEE Transactions on Consumer Electronics, 63(3), 251-257.

[4]. H. M. Soe, & T. M. Naing (2018, May). "Real-Time Hand Pose Recognition Using Faster Region-Based Convolutional Neural Network". In International Conference on Big Data Analysis and Deep Learning Applications (pp. 104-112). Springer, Singapore.

[5]. A. R. Patil, & S. Subbaraman (2018). "A spatiotemporal approach for vision-based hand gesture recognition using Hough transform and neural network." Signal, Image and Video Processing, 1-9.

[6]. X. Yingxin, L. Jinghua, W. Lichun, & K. Dehui (2016, December). "A Robust Hand Gesture Recognition Method via Convolutional Neural Network". In Digital Home (ICDH), 2016 6th International Conference on (pp. 64-67). IEEE.

[7]. H. Y. Lai, H. Y. Ke, & Y. C. Hsu (2018). "Real-time Hand Gesture Recognition System and Application". Sensors and Materials, 30(4), 869-884.

[8]. A. V. Dehankar, S. Jain, & V. M. Thakare (2017, December). "Performance analysis of RTEPI method for real time hand gesture recognition". In 2017 International Conference on Intelligent Sustainable Systems (ICISS) (pp. 1031-1036). IEEE.

[9]. N. H. Dardas, & N. D. Georganas (2011). "Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques". IEEE Transactions on Instrumentation and Measurement, 60(11), 3592-3607.

[10]. T. Chaikhumpha, & P. Chomphuwiset (2018, January). "Real—time two hand gesture recognition with condensation and hidden Markov models". In Advanced Image Technology (IWAIT), 2018 International Workshop on (pp. 1-4). IEEE.

[11]. J. Singha, & R. H. Laskar (2017). "Hand gesture recognition using two-level speed normalization, feature selection and classifier fusion". Multimedia Systems, 23(4), 499-514.

[12]. S. Hussain, R. Saxena, X. Han, J. A. Khan, & H. Shin (2017, November). "Hand gesture recognition using deep learning". In SoC Design Conference (ISOCC), 2017 International (pp. 48-49). IEEE.

[13]. Z. Qiu-yu, L. Jun-chi, Z. Mo-Yi, D. Hong-xiang, & L. Lu (2015). "Hand gesture segmentation method based on YCbCr color space and K-means clustering". Interaction, 8, 106-116.

[14]. A. Choudhury, A. K. Talukdar, & K. K. Sarma (2014, February). "A novel hand segmentation method for multiple-hand gesture recognition system under complex background". In Signal Processing and Integrated Networks (SPIN), 2014 International Conference on (pp. 136-140). IEEE.

[15]. D. S. Alex, & A. WAHI (2014). "BSFD: Background Subtraction Frame Difference Algorithm For Moving Object Detection And Extraction". Journal of Theoretical & Applied Information Technology, 60(3).

[16]. D. Zhang, & G. Lu (2004). "Review of shape representation and description techniques". Pattern recognition, 37(1), 1-19.