A CRITICAL SURVEY ON STATIC AND DYNAMIC HAND GESTURE RECOGNITION

Phyu Myo Thwe(1), Nwe Ni Kyaw(2), Moe Moe Htay(3), G R Sihna(4),Phyu Phyu Khaing(5)
(1)University of Computer Studies (Kalay), Myanmar
(2)(4)(5)Myanmar Institute of Information Technology,Mandalay, Myanmar
(3)University of Computer Studies (Pinlon), Myanmar

Email: phyumyothwe1@gmail.com[1], nwe_ni_kyaw@miit.edu.mm[2], moemoehtay.m2h@gmail.com[3], gr_sinha@miit.edu.mm[4], phyu_phyu_khaing@miit.edu.mm

ABSTRACT

Gestures are the movement of the face, hand, or other parts of the body. So, hand gestures are the movement of hand. Hand gestures give more meaningful information than other types of gestures. Hand gesture recognition is used enormously in recent years to interact with humans and machines. 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 the gesture recognition system is to develop natural device-free interfaces that recognize hand gestures as commands or basic gestural control systems for human-machine interaction and human-robot communication and used them to control electronic devices. This paper reviewed the most commonly used hand gesture recognition 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 the mostly deployed human-computer interactive devices such as a joystick, keyboard, mouse, etc. Hand Gesture can be used in many application areas such as sign language recognition, playing the game, smart home system, controlling the robot, controlling the computer software and many others. There are two types of gestures that are static and dynamic as shown in Fig.1. According to the types of gestures, the processing stages are little changes. In static hand gesture recognition, the general process contains three stages: (i) pre-processing the original input image (ii) feature extraction (iii) recognition the gesture types. In dynamic hand gesture recognition, tracking the hand or transform as frames from the video sequence and then performs feature extraction and then performs feature extraction and finally recognizing the hand gestures. The general hand gestures recognition steps as shown in Fig.2.

Deep learning based method has become popular technologies in recent years. It is proved to very effective especially for computer vision and speech recognition. Among deep learning techniques, CNNs are most appropriate for computer vision tasks.

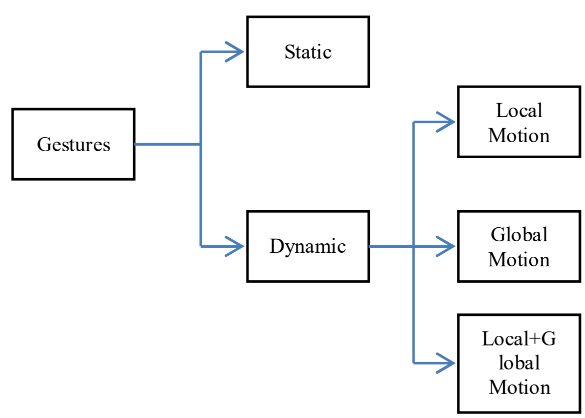

Fig.1. Types of Gestures

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

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.

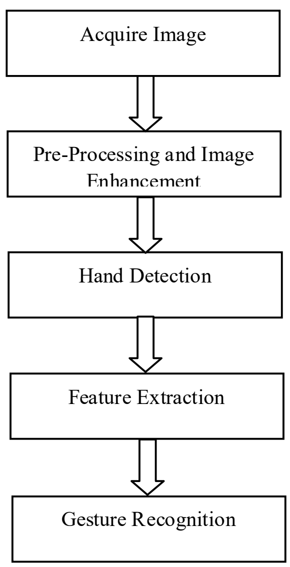

Fig.2. General Hand Gesture Recognition Stages

2. RELATED WORKS OF CURRENT STUDY

Kopuklu et al. (2019) used two models such as detector (ResNet-10) and Classifier (ResNeXt-101) architecture for real time hand gesture recognition system. The detector (ResNet-10) distinguishes between types of gesture and non-gesture classes on the sequence of images. After that, if the detector detects a gesture then the gesture fed into the classifier queues. The classifier (ResNeXt-101) classifies different types of gestures classes. This paper is implemented on two different datasets and classification accuracy 94.04 % and 83.82 % respectively. [1]

Wu X. Y. (2019) proposed double channel convolutional neural network (DC_CNN). In this paper, DC_CNN is experimented on Jochen Triesch Database (JTD) and the NAO Camera hand posture Database (NCD). The author performed pre-processing, denoising and edge detection on the original input images. The accuracy of this research are 98.02 % on JTD database and 97.29 % on NCD database. Double Channel Convolutional Neural Network improved recognition accuracy than traditional single channel convolutional network because DC_CNN can extract wider range of important features. As future work, the training can be carried out with semi-supervised learning or unsupervised learning to improve the system accuracy and to independent on the model [2].

Sahoo et al. (2018) used grey world algorithm to remove illumination variation from the input hand gesture images and discrete wavelet transform and fisher ration (F-ration) used to extract features. Finally, the hand gestures recognized based on Support Vector Machine Algorithm. This paper is experimented on three different datasets such as MUDB dataset, Jochen-Triesch dataset and Self-Constructed dataset. The mean accuracy was obtained 98.64 %, 95.42 % and 99.08 % respectively. The issues of this paper can be misclassifying gestures vocabulary when rotation noise above 15 degree [3].

Lee et al. (2018) implemented static hand gesture recognition using wristband-based contour features. Wristband detection and skin color detection used to detect the hand and watershed segmentation and also used region merging techniques to remove the overlap condition. Finally, Minimum cosine distance used to recognize the static hand gestures. In this paper, the classification accuracy is up to 99.31 % for 29 gestures. The general issue of this paper is difficult to detect the wristband when the background contains dark color and dark shadows within the gesturing hand.[4]

I. COMPARISON OF METHODS

The comparison of the most significant papers and based on the literature review of several papers are given below in table.1.

Table.1 Comparison of Major Important Literature Survey

References	Finding and Limitation
Plouffee et al. (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 et al. (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.
Sahoo et al. (2018)	1. Grey world algorithm was applied to remove illumination changing. After that discrete wavelet transform and F-ration (Fisher ration) also used to extract features. 2. Can be misclassifying the gestures vocabulary when rotation noise above 15 degree. 3. Recognition accuracy of this paper 94.04 % on Ego Gesture dataset and 83.82 on NVIDIA dynamic hand gesture dataset. 4. Different weighting approach for deep structural network will utilize to improve the system performance as a future work.
Wu X. Y. (2019)	1. Studied double channel convolutional neural network (DC_CNN) and obtained more accuracy than traditional single channel convolution neural network (CNN). 2. Recognition accuracy up to 98.02 % on JTD dataset and 97.27 % on NCD dataset. 3. Training is only supervised learning. 4. Semi-supervised learning or unsupervised learning can be carried out to improve the accuracy and to remove the dependence on the model.

3. 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 applied 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.

4. DIFFERENT DATASETS AND EVALUATION MATRICES

A number of datasets are used for training, validation, testing and evaluation for static and dynamic hand gesture recognition methods. There are many various types of datasets with different properties such as static hand posture, dynamic hand gesture, static clutter background, dynamic background, and dynamic illumination condition, under control environment, real world condition and different image size. The popular datasets are NUS Hand Pose Dataset II [5], EgoGesture Dataset [6], Jochen Triesch Database (JTD), Cambridege Hand Gesture Dataset [6], nvGesture Dataset and ChaLearn MMGR Dataset. These datasets are described in section 4 (4.1). In this section, we show some sample images of static and dynamic hand gesture image.

Fig 3. Hand Gesture Images from NUS Hand Pose Dataset II.

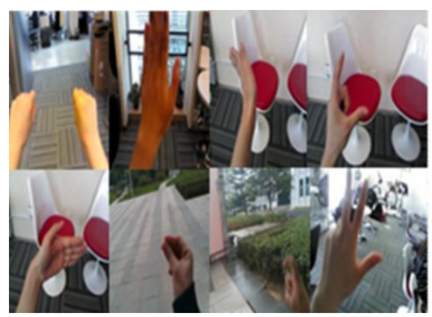

Fig 4. Static and Dynamic Hannd Gesture Images from EgoGesture Dataset.

4.1.Datasets

In this section states the nature of different static and dynamic hand gestures with simplest background, complex background and various lighting conditions.

NUS Hand Post Dataset II [5] is a dataset for hand gesture recognition with complex background condition. NUS hand pose dataset II includes 2000 hand images, 2000 background images and 750 hand images with human noise. 10 different static gestures of hand are defined in this dataset.

EgoGesture Dataset [6] contains 2,081 RGB-D videos, 24,161 gesture samples and 2,953,224 frames from 50 distinct subjects. In this dataset, there are 1239 videos, 14416 gestures sample, 30 subjects for training and 411 videos, 4977 gesture sample for validation, 10 subjects. The performance of the system can be test with 431 videos, 4768 gestures sample, and 10 subjects. The dataset consists of 83 different types of static and dynamic gestures from 6 different scenes.

Jochen Triesch Database (JTD) has 10 gestures classes of 24 people in light condition, dark condition and complex background condition. The size of all images 128*128 and centered in hand gesture.

Cambridege Hand Gesture Dataset consists of 900 video sequences and above 50000 hand gesture images from 2 different scenes. In this dataset, there are 9 classes of dynamic hand gestures with various illumination conditions. But, this dataset is only simple background in under control environment.

nvGesture Dataset has 1532 gestures samples, 25 labels from 20 different subjects. But, this dataset consist only one scene.

ChaLearn MMGR Dataset is a very big hand gesture datasets. This datasets contains 47933 gestures samples and 249 labels from 21 subjects. The dataset can be used for classification and detection task.

Massey University Dataset [14] has 2524 images for 26 poses from five different directions such as top, bottom, left, right and diffuse and various illumination conditions.

Jochen-Triesch Dataset [13] is benchmark dataset. In this dataset contains 10 different alphabets hand gestures. But, the background condition is uniformly.

4.2. Evaluation Metrics

The different kinds of evaluation metrics are used to calculate the quality of hand gesture recognition system. The most commonly used evaluation metrics are discussed in section 4 (4.2). The dataset and evaluation metrics of hand gesture recognition system are listed in table 2.

Levenshtein distance [1] is used as an evaluation metric that is used to evaluate the performance of real-time hand gesture detection and classification system. It can be measure misclassification of the gestures classes, multiple detection, and missing detections of the gesture types at the same time. Confusion Matrix [9, 12] is commonly used table to measure the performance of the classification or recognition accuracy. It is easy to understand but the technology can be confusing.

Table.2 An Summary Of Datasets And Evaluation Metrices

References	**Datasets**	**Evaluation Metrics**
Kopuklu et al. (2019)	EgoGesture,NVIDIA Dynamic Hand Gesture datasets	Levenshtein Distance
Sahoo et al. (2018)	Massey University dataset, Jochen-Triesch dataset, Self-Constructed dataset	Confusion Matrices
Bao et al (2017)	Self-Constructed dataset	Confusion Matrices

4. CONCLUSIONS

Static and Dynamic Hand Gesture Recognition have been used many application areas such as sign language recognition, robot control, playing game with hand gesture and smart home system. In general, hand gesture recognition system contains pre-processing, hand detection, feature extraction and recognition steps. This paper reviewed the current challenges that need to solve as the research issues and how to collect the data for hand gestures and then the nature of hand gesture datasets.

ACKNOWLEDMENT

The authors would like to thank the anonymous reviewers and the editor for their valuable comments and suggestions, which improvement the quality of this paper.

REFERENCES

[1] Köpüklü, O., Gunduz, A., Kose, N., & Rigoll, G. (2019). Real-time Hand Gesture Detection and Classification Using Convolutional Neural Networks. *arXiv preprint arXiv:1901.10323*.

[2] Wu, X. Y. (2019). A hand gesture recognition algorithm based on DC-CNN. *Multimedia Tools and Applications*, 1-13.

[3] Sahoo, J., Ari, S., & Ghosh, D. (2018). Hand Gesture Recognition using DWT and F-ratio Based Feature Descriptor. *IET Image Processing*.

[4] Lee, D. L., & You, W. S. (2017). Recognition of complex static hand gestures by using the wristband-based contour features. *IET Image Processing*, *12*(1), 80-87.

[5] https://www.ece.nus.edu.sg/stfpage/elepv/NUS-HandSet/

[6] Zhang, Y., Cao, C., Cheng, J., & Lu, H. (2018). Ego Gesture: A New Dataset and Benchmark for Egocentric Hand Gesture recognition. *IEEE Transactions on Multimedia*, *20*(5), 1038-1050.

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

[8] Yingxin, X., Jinghua, L., Lichun, W., & Dehui, K. (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.

[9] Sahoo, J., Ari, S., & Ghosh, D. (2018). Hand Gesture Recognition using DWT and F-ratio Based Feature Descriptor. *IET Image Processing*.

[10] Wu, X. Y. (2019). A hand gesture recognition algorithm based on DC-CNN. *Multimedia Tools and Applications*, 1-13.

[11] Köpüklü, O., Gunduz, A., Kose, N., & Rigoll, G. (2019). Real-time Hand Gesture Detection and Classification Using Convolutional Neural Networks. *arXiv preprint arXiv:1901.10323*.

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

[13] https://www.idiap.ch/resource/gestures/

[14] Barczak, A. L. C., Reyes, N. H., Abastillas, M., Piccio, A., & Susnjak, T. (2011). A new 2D static hand gesture colour image dataset for ASL gestures.