Overview of Hand Gesture Recognition Basic Conceptual Framework

Ya-nan XU¹ and Yun-hai DAI²,*

¹Department of Game and Multimedia Engineering, PaiChai University, 155-40 Baejae-ro, Seo-gu, Daejeon, 35345, South Korea
²Department of General Education and International Trade, Namseoul University, 91 Daehak-ro Seonghwan-eup Sebuk-gu Cheonansi Chungcheongnam-do, 31020, South Korea

*Corresponding author

Keywords: Hand gesture, Human-computer interaction, Vision based hand gesture recognition.

Abstract. With the development of science and technology, human-computer interaction is not only an indispensable technology of research projects such as robot control and IOT, but also has gradually penetrated into the daily life of most people. Because of the inconvenience and dependence on the hard devices of traditional human-computer mode, the finding of more natural human-computer interactive ways is particularly urgent. Hand gesture is vivid, intuitive and contains a lot of interactive information, with same expression ability as natural languages such as spoken language and written language, it can act as a means of natural communication between human and machine. This paper offered an overview of hand gesture recognition including its definition and classification, the general framework of vision based hand gesture recognition was then analyzed in detail.

Introduction

Human computer interaction refers to the relation between human and machine, such as controlling a robot or conveying meaningful information. The consideration of usability, the feasibility and efficiency that system can operate and perform a specific user’s purpose, is one of the main characteristics when designing a human computer interaction system [1]. The traditional human-computer interaction mode from the original keyboard to current mouse, joystick and wireless input devices, greatly facilitates the interaction between people and computers and makes it easier for people to operate the computer and improve work efficiency, but these interactions reply on additional input devices, and do not conform to the people’s interaction habits, so this kind of interaction mode cannot completely meet the usability satisfaction of human-computer interaction [2]. Currently, how to make use of the biological characteristics of human beings has becoming a hot focus of natural human-computer interaction mode with the rapid development of machine vision technology, especially image processing and recognition technology. Among the human-computer interaction researches based on biological characteristics such as hand gesture recognition, facial expression recognition, face recognition, lip reading recognition, etc. [3], hand gesture recognition is the most popular one since hand is the most flexible part of human body, it can express rich and various meaning, and more natural, direct compared to other human characteristics, it is an important part of body language, and therefore, can act as a means of natural communication between human and machine.

Hand Gesture Recognition Overview

Hand Gesture Definition

Hand gesture, in a broad sense, is the movement of hand made by people consciously (including the bending and stretching of fingers, rotation of wrist, movement of hands in space), whether it is used to operate a tool or an object, to perform a task or to communicate, hand gesture is always capable of
expressing a signer’s intention. From the point of view of hand gesture recognition, hand gesture is defined as: a variety of gestures or movements produced by hands or arms combined.

**Hand Gesture Classification**

There are multiply classification methods for hand gesture recognition, they are mainly include:

1) Real variable characteristic based
   
   Gesture implementation is a dynamic process, its characteristics are manifested in the changing of hand shape caused by fingers bending, and position or orientation’s changing of hand in space, so the description of hand gesture needs to be considered from two aspects of time and space. According to the real variable characteristic, hand gestures can be divided into static and dynamic ones. A static gesture refers to the space attitude of hand at a certain point of time, a dynamic gesture is a sequence of attitudes in a period of time.

2) Expressive meaning based
   
   Psychologist Quek and Pavlovic proposed the classification of hand gestures based on expressive meaning in 1940s (refer with: Fig. 1). They defined the movement of hand to meaningless action and meaningful action with the conveyance ability of user's intention, and the latter is hand gesture, and it then can be divided into operation and communication ones. Operation gesture dose not express any information but just performs tasks such as grasping and gripping, communication gesture is also called interactive gesture and composed with action gesture and sign gesture. Action gesture refers to hand action such as imitation gesture and pointing gesture, sign gesture such as language used by deaf-mutes, has the function of language description and can be divided into instruction and modal gesture.

3) Input equipment based
   
   Depending on the different ways of gesture image acquisition, hand gesture recognition can be divided into sensor included acceleration hardware based and computer vision based ones.

   Some sensor included acceleration hardware (e.g. Mouse, pen and data gloves) assisted hand gesture recognition has advantages of applicable to general desktop systems with high accuracy and processing speed, but its shortcoming is that only the overall movement of hand can be recognized but not the movement of fingers, that is, only when some movement or orientation change of mouse cursor or nib occur can regard them as gestures.

   Vision based hand gesture recognition uses the camera to collect continuous 2D image sequence, through the extraction and analysis of the image to perform gesture recognition, the biggest advantage of this method is that the input is simple and easy to operate, does not rely on additional devices and interferes. Because of the natural and intuitive interactive way of vision based hand gesture recognition, it becomes the trend and goal of current gesture recognition technology. However, it is difficult to extract the original gesture parts from visual information completely, therefore, its identifiable gesture set is relatively small, has low recognition rate and poor real time performance as
well, but these shortcoming are also the reasons why vision based hand gesture recognition system has become a hot research topic in recent years.

**Vision based Hand Gesture Recognition**

The general flowchart of vision based hand gesture recognition mainly includes (refer with: Fig. 2): sample images capturing, image processing, feature extraction and classification four parts [4].

![Figure 2. Architecture of hand gesture recognition.](image)

**Sample Image Capturing**

Image capturing is to acquire an image (static gesture) or a sequence of images (dynamic gesture), which can be used for processing next. This process is mostly done using a single camera with a frontal view of human hand performing some gestures. There are also systems use multiple cameras in order to get more detailed information of hand gesture [5], this kind of system has the advantage of allowing the recognition of occluded gestures benefited from the scene captured by the other camera from another perspective. Another system with the camera mounted on a hat presented in [6], it is only set to capture the area in front of the hat wearer, its advantage is that the position of camera is always adapted however the people moves or turns around.

**Image Processing**

The basic aim of image processing is to optimally process the sample image, extract its binary image and prepare for the feature extraction next. This process can be mainly divided into color space conversion, noise filtering and threshold segmentation (refer with: Fig. 3).

Segmentation is the most important step of image preprocessing, good image preprocessing leads to perfect features extraction, it divides the input image into regions separated by boundaries. Since human skin is easy and invariant to the changing of scale, translation and rotation, it is commonly used for segmentation [7]. The process of skin color based hand gesture image segmentation generally includes: creating a probabilistic model of skin color to calculate the probability of each pixel, finding the interested coarse regions with comparing with the threshold value, more analysis and filtering being carried out for example involve the size or perimeter of the located regions in order to exclude noisy regions such as faces.
Feature Extraction

The aim of feature extraction is to find and extract features that can be used to determine the meaning of a given gesture, a feature or a set of features should have the abilities to describe the gesture uniquely and be robust to the shift and rotation of hand gesture in order to achieve a reliable recognition. The commonly used features of static hand gestures include: gradient histogram, image subspace projection, shape features, etc. [8]. Traditional gradient histogram is easy to calculate and implement, it has the invariance of translation but not rotation. Image subspace projection method, such as PCA and ICA, is a kind of statistical signal processing technique, which is able to removing the correlation of higher-order statistics and making relatively comprehensive representation of the local features of training sample images, however, the feature invariance of this method is acquired in the training of a large number of samples, once the training samples are not able to cover all the position, scale and rotation angle, it cannot achieve the extraction of invariant features. Shape including contour, silhouette, fingertip, palm center based method has fully considered the invariance of translation, size and rotation in the process of features extraction, has becoming the most commonly used feature extraction algorithm for static gestures recently.

Classification and Recognition

The classification represents the task of assigning the extracted features to some predefined classes in order to recognize the hand gesture. The classification mainly aims to find the best matching reference features of the features extracted between the test and standard images. Before recognition, the system should be trained with enough data (refer with: Fig. 4) so that a new feature vector can be classified with good accuracy [9]. The most commonly used hand gesture recognition techniques include: template matching, artificial neural network (ANN), hidden Markov model (HMM) and Geometric features based recognition, etc.

Conclusion

Hand gesture recognition provides a natural way to interact and communicate with machines for all human dialogs. The ability for computer to visually recognize hand gestures is essential for future human-computer interaction. However, hand gestures recognition is an extremely challenging interdisciplinary project due to reasons of: (1) hand gestures are rich in diversities, multi meanings, and space time varieties; (2) human hand is complex non rigid objects, it has a lot of redundant information; (3) due to the variety of background and unforeseen environment factors, it is always a
difficult problem to divide the human hand area and other background areas in vision based hand gesture recognition system.

In this paper, a comprehensive overview of hand gesture recognition is presented to address the basic concepts and research approaches in each stage. Methods chosen in each stage of the four steps of vision based hand gesture recognition framework can highly affect each other and whole recognition performance.

Acknowledgement
Funding for this paper was provided by Namseoul University.

References
[1] F. Karray, M. Alemzadeh, Human-computer interaction: overview on state of the Art. International Journal on Smart Sensing and Intelligent System, 1 (2008) 137-159.
[2] M. R. Malgireddy, J. J. Corso, A framework for hand gesture recognition and spotting using sub-gesture modeling, in: 20th International Conference on Pattern Recognition, Turkey, 2010, pp. 3780-3782.
[3] N. D. Binh, T. Ejima, Real-time hand gesture recognition using Pseudo 3-D hidden Markov model. In: 5th IEEE International Conference on Cognitive Informatics, China, 2006, pp. 820-824.
[4] B. Tusor, A. R. Varkonyi-Koczy, Circular fuzzy neural network based hand gesture and posture modeling, in: Instrumentation and Measurement Technology Conference, USA, 2010, pp. 815-820.
[5] H. Zhou, D. Lin, Static hand gesture recognition based on local orientation histogram feature distribution model, In: Computer Vision and Pattern Recognition Workshop, USA, 2004, pp. 161-161.
[6] T. Starnet, A. Pentland, Real-time American sign language recognition using desk and wearable computer based video, IEEE Transaction on Pattern Analysis and Machine Intelligence, 20 (1998) 1371-1375.
[7] S. G. Wysoski, M. V. Lamar, A rotation invariant approach on static-gesture recognition using boundary histograms and neural networks, in: Proceedings of the 9th International Conference on Neural Information Processing, Singapore, 2002, pp. 2137-2141.
[8] A. Utsumi, J. Ohya, Image segmentation for human tracking using sequential-image based hierarchical adaptation, in: Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, USA, 1998, pp. 911-916.
[9] Y. Kuno, M. Sakamoto, Vision-based human computer interface with user-centered frame, in: Proceeding of the IEEE/RSJ/GI International Conference on Intelligent Robots and Systems, Germany, 1994, pp. 2023-2029.