Gesture recognition for Human Computer Interaction

Problem Statement

There are many applications where hand gesture can be used for interaction with systems like, videogames, controlling UAV's, medical equipment's, etc. These hand gestures can also be used by handicapped people to interact with the systems. Classical interactions tools like keyboard, mouse, touchscreen, etc. may limit the way we use the system. All these systems require physical contact, in order to interact with system. Gestures can interpret same functionality without physically interacting with the interfacing devices. The problem lies in understanding these gestures, as for different people, the same gesture may look different for performing the same task. This problem may be overthrown by the use of Deep Learning approaches. Convolution neural networks (CNN/'s) are proving to be ultimate tool to process such recognition systems. The only problem with the deep learning approaches it that they may work poorly in real world recognition. High computing power is required in order to process gestures.

Background

There are many hand gesture recognition systems which have been implemented many times. There are two main categories in recognizing hand gestures; one is by use of some hardware device, or by use of captured video of hand gesture. Both the approaches have quite have evolved from the inception of the concepts of gesture recognitions.

Many approaches are implemented for hand gesture recognition using hardware, where the user interact with the hardware and each gesture is recognised as a command to the system. These systems can be seen in many recent applications like Google Glass project by Google, Smart watches with gesture recognition using gyroscopes and accelerometers, laptop touchpads, etc. These applications use many different approaches like; rule-based approaches, machine learning approaches to classify the gestures, etc.

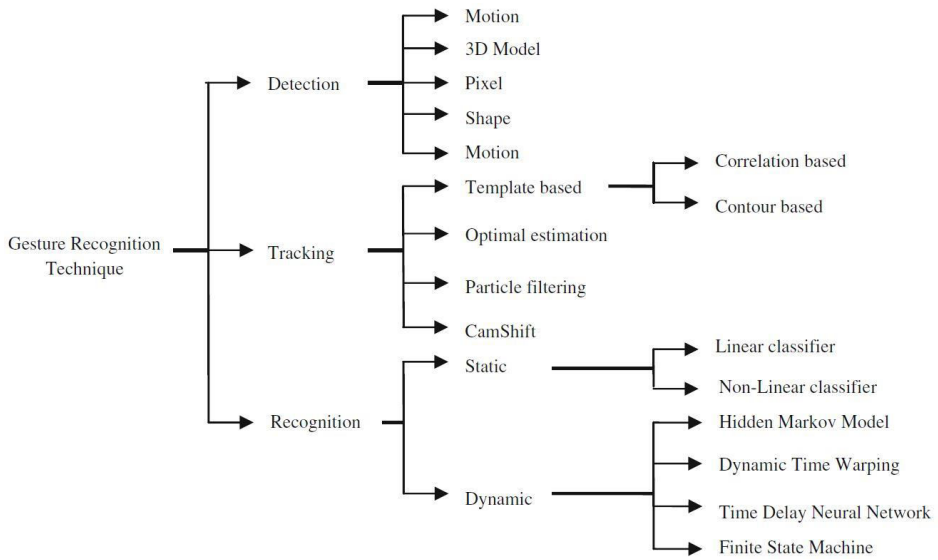

Figure 1: Vision Based Hand Gesture Recognition Techniques ["Vision based hand gesture recognition for human computer interaction: a survey" by Siddharth S. Rautaray and Anupam Agrawal published in Springer Science+Business Media Dordrecht 2012.

Computer vision based hand gesture recognizers use only camera to capture the movement of the hand, and then tries to understand the gesture. Hand recognition and then recognizing what movement is done, as three fundamental steps in computer vision approach; One is detection, another one is tracking and the third one is recognition.

The above figure 1 provides information about several techniques used until 2012 in computer vision for hand gesture recognition. After 2015, many machine learning and deep learning approaches begin to cloud in the field of detecting and recognizing. Let us concentrate in next section how we can use machine learning or deep learning approaches in Hand Gesture Recognition for Human computer interaction.

Methodology

A Recurrent 3D Convolutional Neural Network can be used to design a model for hand gesture recognition. Figure 2 gives the model architecture, which consists of deep 3D-CNN for spatio-temporal feature extraction. A recurrent layer for global temporal modelling. And a softmax layer for predicting the class-conditional gesture probabilities. Finally based on the highest class-conditional probability the gesture is selected and the operating is performed.

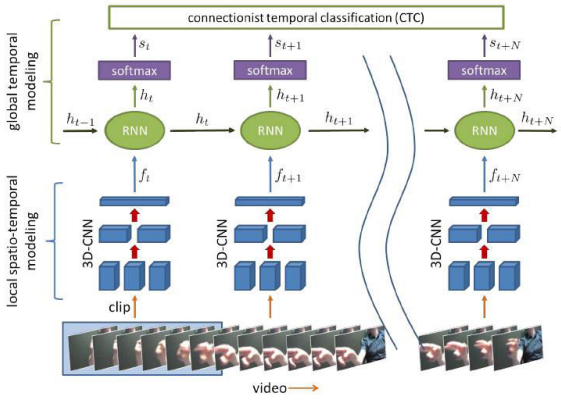

Figure 2: Classification of dynamic gestures with R3DCNN. ["Online Detection and Classification of Dynamic Hand Gestures with Recurrent 3D Convolutional Neural Networks" by Pavlo Molchanov, Xiaodong Yang, Shalini Gupta (NVIDIA), Kihwan Kim, Stephen Tyree, Jan Kautz. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016]

A video clip is provided to the 3D-CNN layer as features by applying spatio-temporal filters to the clip. Here the next layer is Recurrent layer, it acts as the hidden layer whose function is to find the relation between previous frame of the clip and the current frame of the clip. Softmax layer transforms the hidden state output into class-condition probability. Based on this probability a operation is classified.

Experimental Design

Dataset: Dataset used in the above method is provided by NVIDIA , called NVIDIA Hand Gesture Dataset[link] . The data consists of 25 classes for different gestures. It is the largest dataset available for hand gesture recognition.

Evaluation Measures: The model can be evaluated on the basis of accuracy of prediction of class. That is human accuracy is 88.4% in classifying the hand gesture. If the model gets anywhere near to 80 to 85% accuracy then the model is good. The evaluation can also be done on F1 scores, this measurement can help to understand whether the model will work properly in real world scenario.

Software and Hardware Requirements: Python based Computer Vision and Deep Learning libraries will be exploited for the development and experimentation of the project. Tools such as Anaconda Python, and libraries such as OpenCV, Tensorflow, and Keras will be utilized for this process. Training will be conducted on NVIDIA GPUs for training the end-to-end version of CNN based object detection model.