Neural Network Approaches for Visual Recognition of Hand Gestures

Prof. A. V. Deorankar, Neha V. Tavari

Prof. A. V. Deorankar is currently working as an associate professor at Govt. college of Engg. Amravati, India. E-mail: avdeorankar@gmail.com
Neha Tavari is currently pursuing master’s degree program in Computer science and engineering in Govt. college of Engg. Amravati, India. E-mail: tavarineha@gmail.com

KeyWords
Artificial neural network, Computer vision, Deaf and dumb, Hand gesture, Human computer interaction, Sign language.

ABSTRACT
The purpose of gesture recognition researches is to develop a system that can easily identify gestures, and use them for device control, or convey some formations. Hand gestures are the most expressive mode of communication between human and physically impaired people. Gestures are used to form the language which is most commonly used by deaf and dumb community known as sign language. Hand gesture is an active area of research in the computer vision, mainly for the purpose of sign language recognition and human computer interaction. This paper explores Artificial Neural Networks based approaches in the area of hand gesture recognition. Several hand gesture recognition researches that makes the use of neural networks are discussed in this paper.
1. INTRODUCTION

The expectation of widely extensive range of computer systems with the rapid development of information technology in our life, would be inter in our environments. These environments should make simple, natural and easy to use interfaces for human computer-interaction (HCI). The user interface of any personal computer has evolved from primitive text user interfaces to graphical user interfaces (GUIs) which still limited to keyboard and mouse input, however, they are inconvenient, unnatural, and not suitable for working in virtual environments. With the use of the hand gestures an efficient alternative would be provided to these onerous interface devices for human-computer interaction [1].

These gestures are used by physically impaired peoples to express their thoughts in terms sign language. Sign Language is a language which uses hand gestures, facial expressions and body movements for communication. Usually deaf people seek the help of sign language interpreters for translating their thoughts to normal people and vice versa. But this system is very costly and does not work throughout the life period of a deaf person. So a system that automatically recognizes the sign language gestures is necessary. Such a system can minimize the gap between deaf people and normal people in the society.[2] Gesture recognition, needs a good interpretation of the hand movement as effectively meaningful commands. For human computer interaction (HCI) interpretation system there are two commonly approaches [1]:

A. DATA GLOVES APPROACHES

These methods employ mechanical or optical sensors attached to a glove that transforms finger flexions into electrical signals to determine the hand posture. Using this method the data is collected by one or more data-glove instruments which have different measures for the joint angles of the hand and degree of freedom (DOF) that contain data position and orientation of the hand used for tracking the hand. However, this method requires the glove must be worn and a wearisome device with a load of cables connected to the computer, which will hampers the naturalness of user-computer interaction.

B. VISION BASED APPROACHES

These techniques based on the how person realize information about the environment. These methods usually done by capturing the input image using camera(s)[1]. For ex. in vision based hand gesture recognition system, the movement of the hand is recorded by video camera(s)[3]. In order to create the database for gesture system, the gestures should be selected with their relevant meaning and each gesture may contain multi samples for increasing the accuracy of the system. In this work we used vision based approaches

![Fig. 1. (a) Data glove (b) Vision based](image)

Vision Based hand gesture recognition approaches can be categories into: appearance based approaches, and 3D model based approaches as follows:

i) Appearance Based Approaches: these approaches use features extracted from visual appearance of the input image model the hand, comparing these modeled features with features extracted from input camera(s) or video input.

ii) 3D Model Based Approaches: Model based approaches depends on the kinematic hand DOF’s of the hand. These methods try to infer some hand parameters like, pose of palm, joint angles from the input image, and make 2D projection from 3D hand model[1].

2. ARTIFICIAL NEURAL NETWORKS: OVERVIEW

During the development through the years the computational variation has growth to new technologies, Artificial Neural Networks are one of the technologies that solved a broad range of problems in an easy and convenient manner. The working concept of Artificial Neural Networks (ANNs) is similar to human nervous system, hence it has synonym with the word neural networks[1]

An artificial neural network involves a network of simple processing elements (artificial neurons) which can exhibit complex global behavior, determined by the connections between the processing elements and element parameters. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. The utility of artificial neural network models lies in the fact that they can be used to infer a function from observations. This is
particularly useful in applications where the complexity of the data or task makes the design of such a function by hand impractical. The tasks to which artificial neural networks are applied. Classification, including pattern and sequence recognition, novelty detection and sequential decision making[4].

A. FEED FORWARD NETWORKS

Feed forward Networks are the simplest devised type of artificial neural network. From its name ‘forward’ the information moves in one direction, starts from the input nodes to the output nodes goes through the hidden nodes (if any) with no cycles, It can be formed with different types of units.

![Simple Feed forward Networks](image)

B. FEED BACKWARD NETWORKS OR RECURRENT NEURAL NETWORK

A recurrent neural network (RNN) is a class of neural network where connections between units form a directed cycle. This creates an internal state of the network which allows it to exhibit dynamic temporal behavior. Recurrent neural network can be models with bi-directional data flow, which allows connection loops between perceptron. Unlike feedforward neural networks, RNNs can use their internal memory to process arbitrary sequences of inputs. In fully connected network there are no distinct input layers of nodes , and each node has input from all other nodes, feedback to the node itself is possible.

![An example of a fully connected recurrent neural network](image)

C. KOHONEN SELF-ORGANIZING MAPS (SOM)

Self-Organizing Map is a type of neural network, developed in 1982 by Tuevo Kohonen . ‘Self-Organizing’ called so since no supervision is required and learning by means of unsupervised competitive learning. ‘Maps’ called so since they map the weights to be correspond to the given input, and the nodes in a SOM try to like the inputs presented to them. This is how they learn, can also call as “Feature Maps”. Some of SOM applications are, Color Classification, Image Classification[5]. A self-organizing map consists of components called nodes or neurons. Associated with each node is a weight vector of the same dimension as the input data vectors and a position in the map space. The self-organizing map describes a mapping from a higher dimensional input space to a lower dimensional map space. The procedure for placing a vector from data space onto the map is to find the node with the closest (smallest distance metric) weight vector to the data space vector.
3. **GESTURE RECOGNITION USING ARTIFICIAL NEURAL NETWORKS**

Because of Artificial Neural Network ANNs nature that consist of many interconnected processing elements, it can be constructed for problems as mentioned in; searching for identification and control, game-playing and decision making, pattern recognition medical diagnosis, financial applications, and data mining. The most commonly used algorithm for training a feedforward neural network is the backpropagation algorithm. It works by the principle of “backward propagation of errors”. Backpropagation is a supervised learning technique and the network is provided with the pairs of inputs and outputs that the network has to compute. The input patterns are given to the network through the neurons in the input layer and the output of the network is obtained through the neurons in the output layer. Then the backpropagation algorithm computes the difference between actual and expected results and this error value is propagated backwards. The backpropagation algorithm tries to minimize this error until the neural network learns the training data[2]. In our proposed approach, the neural network is trained to classify 36 hand signs. The training dataset contains 360 images with 10 images of each of the 36 signs. In testing phase, a dataset containing 180 images corresponding to 5 images of 36 signs each is used to test the proposed system.

Deepika Tewari, Sanjay Kumar Srivastava proposed an algorithm for hand gesture recognition system in ISL which is based on vision-based approach. An intensity (grayscale) representation of the segmented image is used for further processing. This grayscale version, also called a “skin map,” contains intensity values for skin pixels and the background is represented as black. Then, the Two-Dimensional Discrete Cosine Transform (2D-DCT) for each region is computed, and feature vectors are formed from the DCT coefficients. The DCT can be extended to the transformation of 2D signals or images. This can be achieved in two steps:

i) computing the 1D-DCT of each of the individual rows of the two dimensional image,

ii) computing the 1D-DCT of each column of the image.

DCT-based feature vectors are classified to check whether sign mentioned in the input image is “present” or “not present” in the ISL database using self-organizing map (SOM)[3] with unsupervised learning technique in Artificial Neural Network (ANN). As SOM is based on unsupervised learning, no mediation is needed during the learning and little need to be known about the characteristics of the input data which makes it to be used for clustering data without knowing the class memberships of the input data. The SOM is also known as SOFM, the Self-Organizing Feature Map (SOFM) as it can be used to detect features belonging to the problem. The particular kind of SOM known as a Kohonen Network is used which have feed-forward structure with a single computational layer arranged in rows and columns.
G.R.S. Murthy and R.S. Jadon [7] proposed a model, based on pattern recognition techniques using supervised feed-forward neural net training and back propagation algorithm for classifying hand gestures into ten categories: hand pointing up, pointing down, pointing left, pointing right and pointing front and also to count the number of fingers user was showing. We have applied a simple pattern recognition technique to the problem of hand gesture recognition. This method has a training phase and a testing phase. In the training phase, the user shows hand gestures which were captured using Image Acquisition Toolbox of MATLAB 7.01 and USB based Fronttech e-cam camera.

A novel technique is proposed by Dipak Kumar Ghosh, Samit Ari [8] to obtain a rotation invariant gesture image which coincides the 1st principal component of the segmented hand gestures with vertical axes. The shape of the contour is an important property that can be used to distinguish of the static hand gestures from one class to another. The localized contour sequence (LCS), which has been confirmed to be a very efficient representation of contours, is selected as a feature set of the hand gesture. A contour tracking algorithm is proposed to track the contour of a gesture in the clockwise direction and the contour pixels are numbered sequentially starting from the topmost left contour pixel. After successfully extracting a normalized LCS feature vector of the static hand gesture, the classification job is done via k-mean based radial basis function neural network (RBFNN). Radial basis function neural network (RBFNN) is widely used in pattern recognition tasks for its fast learning algorithms. The centers and spread factor of the radial basis function are important parameter of RBFNN. Several methods have been used to find the centers of the RBFNN. In this paper, a k-means clustering based approach is proposed to determine the centers of the RBFNN.

4. CONCLUSION

In this paper we have presented an idea of hand gesture recognition and Neural Networks approaches. One of the most effective of software computing techniques is Artificial Neural Networks that has many applications on hand gesture recognition problem. Neural Networks system can be applied for extracted features from the input image gestures after applying segmentation, as in to extract the shape of the hand. The methods are implemented completely by utilizing digital image processing techniques so the user does not have to wear any special hardware device to get the features of the hand shape.

REFERENCES


