Implementation and Performance Analysis of Different Hand Gesture Recognition Methods

Md. Manik Ahmed[1], Md. Anwar Hossain[2], A F M Zainul Abadin[3] and A F M Zainul Abadin[4]

[1] Rabindra Maitree University, Kushtia, Bangladesh.

Received: 12 December 2018 Accepted: 3 January 2019 Published: 15 January 2019

Abstract

In recent few years, hand gesture recognition is one of the advanced grooming technologies in the era of human-computer interaction and computer vision due to a wide area of application in the real world. But it is a very complicated task to recognize hand gesture easily due to gesture orientation, light condition, complex background, translation and scaling of gesture images. To remove this limitation, several research works have developed which is successfully decrease this complexity. However, the intention of this paper is proposed and compared four different hand gesture recognition system and apply some optimization technique on it which ridiculously increased the existing model accuracy and model running time. After employed the optimization tricks, the adjusted gesture recognition model accuracy was 93.21

Index terms — convolutional neural network, deep learning, hand gesture recognition, PCA.

1 Introduction

n natural communication with another person, hand gesture recognition plays a vital role to interact with them naturally, convey rich and meaningful information in various ways. Because Gestures are one of the general forms of communication when people from different languages meet, and no one knows in which language they should express their feelings [1]. In compare to other body parts, a human hand which has been treated as a natural organ for a human to human interaction has been used widely for gesturing and can be best suitable for communication between humans and computers [4]. Recently a computer is an essential machine in our society which accomplishes our daily tasks. Human-computer interaction (HCI) is not only the keyboard, mouse interaction but also interaction computer with the human-like gesture, natural language, emotion and body expressions, etc. [2].

For example, if we consider today's world without a computer, then we can easily realize HCI in our society. It is the most critical issue of advanced technology to recognize, classify and interpret various simple hand gestures and apply them in a wide range of application through HCI and computer vision.

In early, there were many gesture recognition techniques have been developed for recognizing and tracking various hand gesture images. Previously developed available hand gesture recognition techniques are instrumented gloves, optical markers and some advanced methods based on image features, color-based, vision-based, depth-based models but have their advantages and limitation [2]. But the previous gesture recognition technique fails to obtain the satisfiable result. Some automatic feature extraction based hand gesture recognition techniques have developed which was remove the limitation of previous work and made a revolutionary change on HCI and computer vision era.

In this paper provides four different hand gesture recognition techniques and compare performance among machine learning based semiautomatic and deep neural network based automatic system. These four methods tested on the same testing data with the same epoch. The classification task usually depends on the two-factor one is time, and another is recognition accuracy. A recognition method will perfect if the method takes low time and high accuracy during the train and test dataset. In this paper provides principal component analysis with backpropagation neural network based model but it showed very poor accuracy close to 77.86%. To remove this limitation presented a deep learning based primary two-layer convolutional neural network (PCNN) model that achieves the accuracy of 91.07% but consumes more time than the previous modes. Then proposed an adjustable CNN model (ACNN) which is similar to the PCNN model but there applied batch normalization and regularization that provide higher accuracy and minimize the running time half of the PCNN model. Finally, we applied two extra CNN layers with the PCNN model but it takes long running time, but provides 96.43% accuracy. There are several typical applications of hand gesture recognition such as virtual game controller, sign language recognition, Directional indication through pointing, making young children to interact with the computer, human-computer interaction, robot control, lie detection, home appliance, Camera control, entertainment, and medical systems, Gesture talk, etc. [3].

The paper has organized as follows: Section 2 gives some literature review of present and previous work. Section 3 describes the methods of the proposed system. Experiment results and discussion have shown in Section 4. Finally, Section 5 concludes this paper.

2 II.

3 Literature Review

Recognition of hand gesture offers a new era and plays a vital role in nonverbal communication and interact with the machine naturally. There is various bodily motion which can originate gesture, but the general form of gesture origination comes from the face and hands. The entire procedure of tracking gestures to their representation and converting them to some purposeful command is known as gesture recognition [4]. The hand gesture is the easiest and potential research area of machine learning and computer vision [5]. In a few years, several methods have been proposed to recognize hand gesture with the adaptive manner. The authors [6] [8] proposed a vision-based hand gesture recognition system that based on skin color model and thresholding approach which is segmented by the skin color model in YCbCr color space and separate hand region from the background by the Otsu thresholding method. Finally, they developed a template-based matching technique by Principal Component Analysis (PCA) for recognition. Their experiment tested on 80 images achieved 91.25% average accuracy and 30 images with 91.43% accuracy on the independent database.

Flores [12] proposed an approach to recognize static hand gestures whose features varied in scale, rotation, translation, illumination, noise, and background. Their approach included applying various digital image processing filter techniques to reduce noise, to improve the contrast under a variant illumination. They separate the hand from the background of the image and finally, and to detect and cut the region containing the hand gesture. Their approach achieved 96.20% recognition accuracy. Bui T.T.T [7] proposed a novel algorithm for face and hand gesture recognition system based on wavelet transform and principal component analysis that processes with two stages. At this stage, they extract object features using wavelet transformation and save it to the database so that they can compare this feature with PCA based extracted feature through the result. They achieved an efficient performance of face recognition (98.40 %) on 7320 testing face images and hand gesture recognition (94.63%) on digital image Nasser H. Dardas and Shreyashi Narayan Sawant [9][10] proposed traditional PCA algorithm for gesture recognition which hand feature or train weight was extracted by projecting each training image onto the most eigenvectors then the small image that contains the detected hand gesture is projected onto the most eigenvectors of training images to form its test weights. Finally, they utilized the euclidian distance to recognize the hand gesture.

The above recognition method doesn't provide sustainable and remarkable results due to their limited accuracy and high time consumption and semiautomatic behavior. Nowadays, Deep learning based Convolutional Neural Networks (CNN) have shown substantial performance in different recognition tasks on computer vision that extend the traditional artificial neural network by adding additional constraints to the earlier layers and increased the depth of the network. A. Krizhevsky [11] proposed an ImageNet Large Scale Visual Recognition Challenge work which is mainly focused on the architecture to achieve great performance on a large number of data set during training.

Gongfa Li [12] proposed a convolutional neural network that removes the traditional feature extraction method and reduces the number of training parameter. They utilized the error backpropagation algorithm for learning the network in an unsupervised way. Finally, they added the support vector machine act as a classifier to improve the validity and robustness of the whole classification function of the convolution neural network. They achieved an efficient performance of gesture recognition average 98.52 % on 7320 gesture images of 10 different people. Yingxin [13] proposed an approach for hand gesture recognition based on the Adapted Deep convolutional neural network (ADCNN) with regularization technique which took shifted and rotated version of hand gesture images that extend the 20% of the original image dimension randomly. Their experiments conducted with a regularization technique on 3750 hand gesture images that remove the overfitting. Their result revealed the ADCNN approach achieved higher recognition accuracy of 99.73% and 4% improvement over the traditional CNN model. Guillaume Devineau [14] proposed an approach using a 3D deep convolutional neural network(3DCNN) for hand gesture recognition using only hand-skeletal data without depth image information. Their proposed 3DCNN only processed sequences of hand-skeletal joints' positions by parallel convolutions. Their experiment achieved a 91.28% classification accuracy for the 14 gesture classes case and an 84.35% classification accuracy for the 28 gesture classes case. In Pei Xu [15] Proposed a hand gesture classification method which is the modified CNN version from LENET 5 using only one cheap monocular camera. Their experiment also introduced the Kalman filter to estimate the hand position based on which the mouse cursor control had realized stably and smoothly. But their implemented system only supported static gesture and worked on 3200 gesture images.

4 III.

5 Methods and Methodologies

This section provides literature behind the Principal component analysis (PCA), a description of Backpropagation neural network (BPNN), a brief discussion of primary 2-layer convolutional neural network (PCNN) architecture and the basic overview of adjusted convolutional neural network (ACNN) which is the optimized version of PCNN, Description of Primary 4-layer CNN to gain better performance. Also, this section provides the Layer operation and configuration table of all neural network that has proposed in this paper.

6 a) PCA Based BPNN Architecture

Principal component analysis (PCA) is a dimensionality reduction technique based on extracting the relevant information of gesture images which is multidimensional. The main objectives of PCA in gesture recognition techniques are data dimension reduction and feature selection to train the Multilayer BPNN. The gesture recognition using PCA based BPNN architecture involves two phases: i) Feature Extraction Phase ii) Classifier Phase. During feature extraction Phase PCA is to reduce the dimensionality of the gesture images while retaining as much information as possible in the original gesture images. Each hand gesture images in the database concatenate form into one matrix. Then PCA is to move the origin to mean of the data by averaging all column matrix database images divided by the total number of hand gesture images. Next, find the normalized images by subtracting the computed average image from each image in database form into a mean centered data matrix. These adjusted images determine how each of the gesture images in the database differs from the average image which has calculated previously. Next, the PCA algorithm calculates and finds the eigenvectors using covariances matrix of normalized hand gesture images that speed up the technique and reduce the number of parameters. Eigenvectors with low eigenvalues that contribute little information in the data representation. In this step also data reduction technique is achieved by truncating the eigenvectors with small eigenvalues. Then this eigenvectors matrix multiplied by each of the normalized gesture vectors to obtain their corresponding gesture space projection. Finally, each image in normalized matrix multiplied with gesture space and created a new gesture descriptor or strong weight that is ready to feed as the input of BPNN.

In the classifier phase, A Backpropagation neural network has used as a classifier which is reverse propagates the error and adjusts the weights to near the target output. In particular, the internal (hidden) layers of multilayer networks learn to represent the intermediate features that are useful for learning the target function and that are only implicit in the network inputs. The classification of hand gesture images involves in two stages one is training stage another is the testing stage. During the training stage, BPNN design is composed of two hidden layers, Input layer, and output layer. The model of BPNN has described in (Table 01). In this stage, the gesture feature vectors that belong to the same classes have used as positive examples, i.e. network gives "1" as output, and negative examples for the others network, i.e. Network gives "0" as output which is the target value. The algorithm used to train the network is the Backpropagation Algorithm. The general idea with the Backpropagation algorithm is to use gradient descent to update the weights to minimize the squared error between the network output values and the target output values. Then, each weight is adjustedusing gradient descentaccording to its contribution to the error. The activation or transfer function used in the Back Propagation neural network is the sigmoid function which maps the output 0 to 1. When the neural network met the stopping condition, then it stops and gives the training output. During the testing stage, it is necessary to extract the feature of all unknown hand gesture images. Then calculate the projection of the test gesture to project the gesture on gesture space and form into a new descriptor. These new descriptors have inputted to every network, and the networks are simulated with these descriptors. The network outputs have compared. If the maximum output exceeds the predefined threshold level, then these new gesture images have decided to belong to the same class gesture with this maximum output. Finally, this architecture calculates overall accuracy depend on correctly recognize of hand gesture images out of total images. ( ) D b) Convolutional Neural Network

In the field of image recognition and computer vision, convolutional neural network (CNN) has achieved the most remarkable results [16]. Recently, with the development of hardware, much research about object recognition using CNN becomes practical and achieves success **??**11] [17]. CNN usually learn features directly from input data and often provides a better classification result in the case where features are hard to be extracted directly, such as image classification.

Traditional CNN contains two parts, namely the feature extraction part also called the hidden layer and the classification part. In feature extraction part the CNN can perform a series of convolutions and pooling operationsduring which the features have detected. In the case of a CNN, the convolution has performed by sliding the filter or kernel over the input images with some stride which the size of the step the convolution filter moves each time and the sum of the convolution produces a feature map. The most common thing in CNN network is to add a pooling layer in between CNN layers which function is to continuously reduce the dimensionality to reduce the number of parameters and computation in the network model. In CNN, two types of pooling layer available are average, and another is maxpool which are reduce the training time and controls overfitting [11]. In the classification part which contains some fully connected layers that can only accept onedimensional data. The output of this layer produces the desired class using an activation function and classifies given images.

7 c) Primary CNN Architecture (PCNN)

In this section of this paper provides the primary structure of deep CNN which contains two convolutional layers, two pooling layers and with two fully connected layers with rectified linear unit (RELU) and one softmax activation function. The structure of the PCNN has shown in Figure 02 (Fig. 2). In this architecture, it also contains a dropout procedure after the first convolution layer and the second dropout has performed after the first fully connected layer. The overall objectives of this primary deep CNN network have automatically extracted the feature from the direct gesture image as input that removes the traditional machine learning semiautomatic and manual techniques.

In PCNN, the first layer starts with a convolutional layer that has a kernel size of 3×3 pixels and contains 32 feature maps that followed by traditional RELU activation function with no padding. This layer performed by fed the hand gesture images with 64×64×3 pixel values and extracts the deeper information for every image in the database. The next layer is a Dropout also called regularization layer that was configured to randomly remove 20 percent of neurons to reduce overfitting or underfitting. The next deeper layer is another convolutional layer that has a kernel size of 3×3 pixels and also contains 32 feature maps followed by a RELU activation function that is the same as the previous layer. The next layer is a MaximumPooling layer that has designed with a pool dimension of 2×2 pixels. The MaxPoolinglayer uses the maximum value to progressively deduct the negative of the representation to reduce the number of parameters and computation in the network, and hence to also control overfitting. The MaxPooling layer extracts subregions of 2×2 of the feature map, keeps their maximum value and discards all other minimum values. A layer called Flatten that converts the two-dimensional image matrix data to a vector, hence allowing the final output to be processed by standard fully connected layers to obtain the next layers. The first fully connected layer in this PCNN with the RELU activation function contains 512 neurons. This layer is followed by a dropout layer toexclude 20% of neurons to reduce overfitting. The final part of the CNN structure is the output layer and acts as aclassifier which is mapped by a Softmax activation function, and contains six neurons represents every class of gesture image. The architecture of the PCNN network has shown in (Figure 02). In this section, provides the fine regulated design of PCNN with a simple modification to get better accuracy and low running time for the real-world applications. There are existed several techniques to optimize the CNN. CNN was parameter sensitive neural network where accuracy and running time largely depend on network parameters. There are several strategies to extend recognition accuracy and running time. In this paper, the overall performance of ACNN was improved by a simple modification of the network parameter that includes i) Downsample, ii) Batch normalization and Kernel size and iii) Regularization. The ACNN is the modified version of PCNN that contains some optimization technique compare with another CNN model shown in (Table ??2): Downsample: In this ACNN architecture provides two extra downsample operation after two convolutional layers with 1×1 kernel size through pooling layer that reduces the feature map dimensionality for computational efficiency, which can, in turn, improve actual performance accuracy.

8 Kernel Size and Batch normalization:

To preserve the low running time it will be necessary to replace filter or kernel size from a higher value to lower. In the very initial step, an acceptable kernel of suitable dimensions is decided to convolve over the input image and identify key features in the images. A larger size kernel can disregard the features and could skip the essential details in the images whereas a smaller size kernel could provide more information leading to more confusion. This ACNN model contains the second convolutional layer that has a kernel size of 2×2 pixels and also holds 64 feature maps followed by a RELU activation function which differs from PCNN. Batch normalization isthe another key factor that speeds up the model run time. This model batch size is 32 that means 32 samples from the training dataset will be used to estimate the error gradient before the model weights are updated.

Regularization: This Deep ACNN deals with a large number of parameters while training the model leads to overfitting. Regularization is the technique that decreases the complexity by constructing the model structure as simple as possible. This model contains two regularizations after the first convolutional layer and the first fully connected layer with 20% and 27% dropout.

9 e) Larger CNN Architecture (LCNN)

This section proposed the Larger CNN (LCNN) architecture that contains four convolutional layers with 3×3 pixels and each convolutional layer followed by RELU activation function, max-pooling layer with 2×2 pool size and batch normalization. The first convolutional layer only contains 32 feature maps, and the rest of the three convolutional layers contains 64,64 and 96 feature maps accordingly. One flattens layer that converts image to vector. Also, this structure presents two fully connected layers with RELU and Softmax activation function and output layer gives the probability value of the recognized class. This architecture did not use any regularization technique to reduce model complexity.

10 IV.

11 Experimental Results and Discussions

This section presents the basic description of Hand gesture Datasets which was used to perform the experiments and fed into PCA with Backpropagation, PCNN, LCNN model and adapted version of the PCNN model called ACNN. Besides this section provides the evolution parameters and performance among these four models. Then it discusses the accuracy and running time and evaluates and measures the best network performance among them.

The presented four architecture are trained and tested on hand gesture dataset taken from the deeplearning.ai (DLAI) **??**17] dataset which contains 1080 randomly organized hand gesture datasets. This dataset contains six unique class from 0 to 5 that depends on the corresponding hand finger. The dimension of the individual image is 64×64×3 where 64 represents the width, the height of the corresponding images and 3 represents the channeli.e., the hand gesture data used in this experiment are color images. We divided the datasets 800 images for training and 280 images for testing from 1080 individual image datasets as shown in (Table **??**2).

12 Conclusion

In this paper, we presented and compared four semi-automatic and automatic gesture recognition methods that classify hand gesture data from a large number of datasets. The semi-automatic method works in a two way. First, it extracts feature from given datasets and feeds it to the classifier for recognition. We In the future work, we would like to apply different optimization technique on the LCNN network so that we can speed up the model run time and the proposed system apply to many applications such as home appliance, Camera control, entertainment, and medical systems, Gesture talk, etc. [1] [2]

1

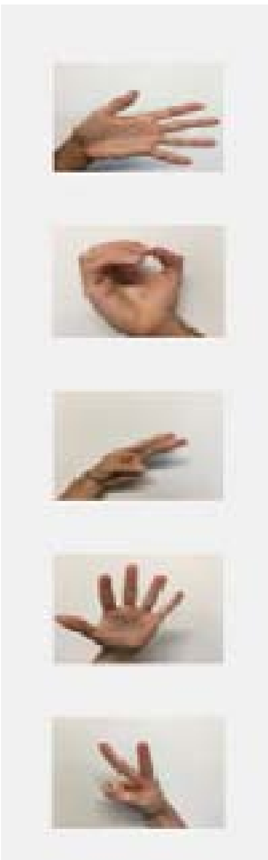


Figure 1: Fig. 1 :

Figure 2: .

2

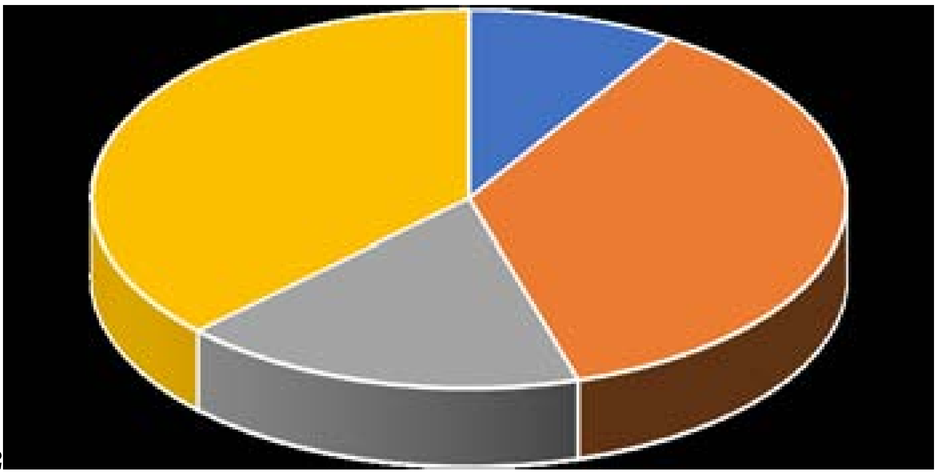
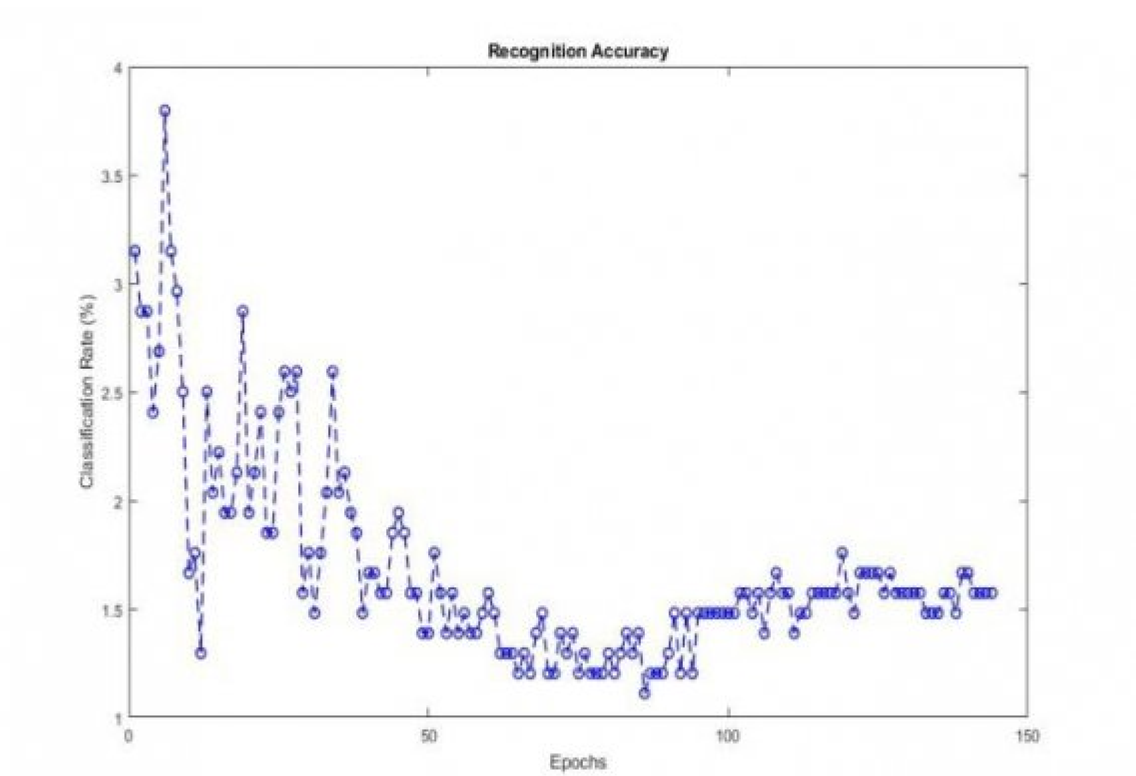
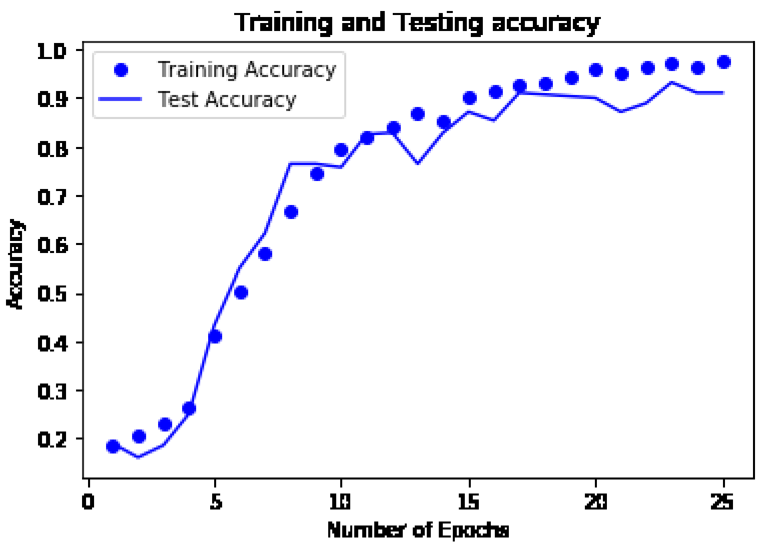


Figure 3: Fig. 2 :



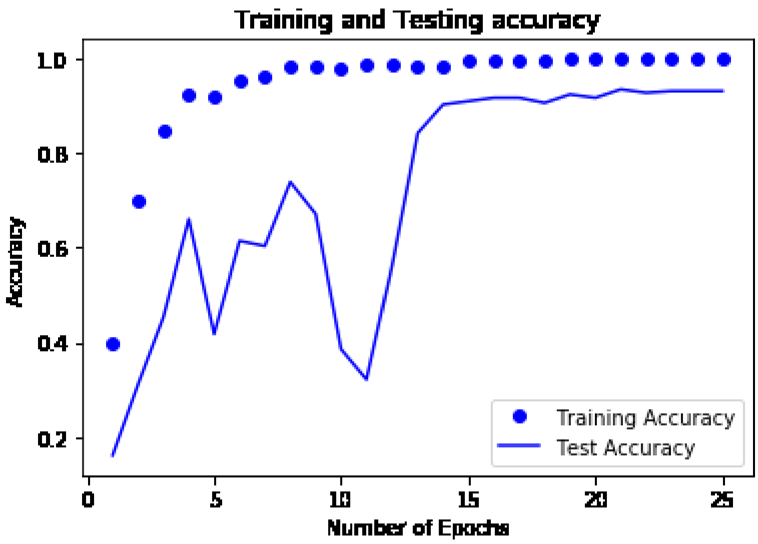
26033

Figure 4: Table 2 : 6 Figure 03 (Fig. 3 :



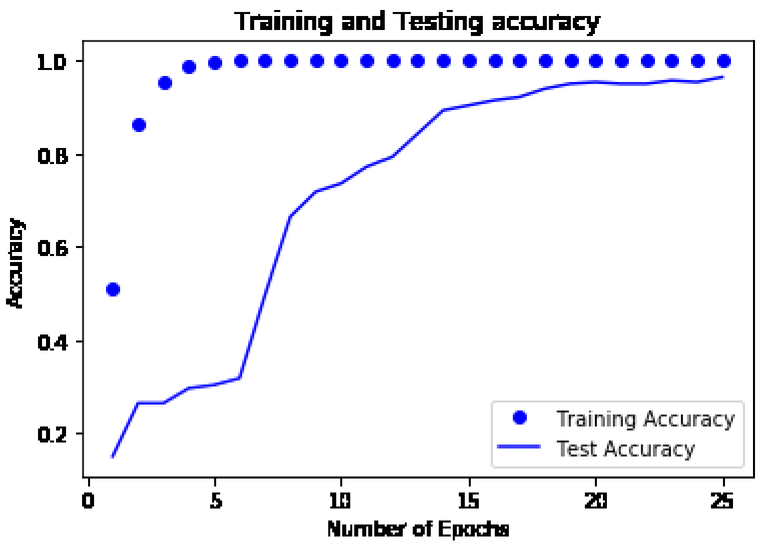
4

Figure 5: Fig. 4 :



5

Figure 6: Fig. 5 :



6

Figure 7: Fig. 6 :

1

Parameter types	Parameter
Layer	4
Input layer neuron	1080
Hidden layer neuron	75 (H1) 50 (H2)
Transfer function	Logsig
Epochs	1000
Learning rule	Backpropagation based gradient descent

Figure 8: Table 1 :

Figure 9:

3

Network	PCA with BPNN	PCNN	ACNN	LCNN
Training Images	800	800	800	800
Testing Images	280	280	280	280
Convolutional layer/hidden layer	2	2	2	4
Learning rate	0.01	0.01	0.01	0.01
Epochs	146	25	25	25
Activation function	Sigmoid	RELU-Softmax	RELU-Softmax	RELU-Softmax
Cost function	Mse	categorical_cros sentropy	categorical_cros sentropy	categorical_cros sentropy
Optimization	SGD	SGD	SGD	SGD
Batch normalization	None	None	Yes	Yes
Regularization	None	Yes	Yes	None
Downsampling	None	Limited	Yes	Limited

Figure 10: Table 3 :

[1]( ) D © 2019 Global Journals Implementation and Performance Analysis of Different Hand Gesture Recognition Methods

[2]© 2019 Global JournalsImplementation and Performance Analysis of Different Hand Gesture Recognition Methods

[Mandeep Kaur Ahuja and Singh (2015)] , Amardeep Mandeep Kaur Ahuja , Singh . *Hand Gesture Recognition Using PCA, IJCSET* July 2015. 5 p. .

[Razavian et al. ()] 'CNN features off-the-shelf: an astounding baseline for recognition'. A S Razavian , H Azizpour , J Sullivan . `https://www` *IEEE conference on computer vision and pattern recognition workshops*, 2014. IEEE. p. .

[Devineau1 and Xi ()] 'Deep Learning for Hand Gesture Recognition on Skeletal Data'. Guillaume Devineau1 , Wang Xi . *13th International Conference on Automatic Face & Gesture Recognition*, 2018. IEEE.

[Bui et al. ()] 'Face and Hand Gesture Recognition Algorithm Based on Wavelet transforms and Principal Component Analysis'. T T T Bui , N H Phan , V G Spitsyn . *7th International Forum on Strategic Technology (IFOST)*, 2012. 2012.

[Nasser et al. (2011)] 'Hand Gesture Detection and Recognition Using Principal Component Analysis'. H Nasser , Emil M Dardas , Petriu . *IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA) Proceedings*, 2011. Sept. 2011.

[Pradipa et al. (2014)] 'Hand Gesture Recognition Analysis of Various Techniques, Methods and Their Algorithms'. R Pradipa , Mss , Kavitha . *International Journal of Innovative Research in Science, Engineering and Technology* March 2014. 3 (3) .

[Li and Tang (2017)] *Hand gesture recognition based on convolution neural network*, Gongfa Li , Heng Tang . 29 december. 2017. Springer transaction.

[Alani et al. ()] 'Hand Gesture Recognition Using an Adapted Convolutional Neural Network with Data Augmentation'. Ali A Alani , Georgina Cosma , T Aboozartaherkhani , Mcginnity . *International Conference on Information Management* 2018. IEEE.

[Sharma and Verma ()] 'Human Computer Interaction using Hand Gesture'. Ram Pratap Sharma , Gyanendra K Verma . *Eleventh International Multi-Conference on Information Processing-2015*, 2015. Elsevier. 54 p. .

[Krizhevsky et al. ()] 'Imagenet classification with deep convolutional neural networks'. A Krizhevsky , I Sutskever , G Hinton . *NIPS*, 2012.

[Panwar and Singh Mehra] Meenakshi Panwar , Pawan Singh Mehra . *Hand Gesture Recognition for Human Computer Interaction, International Conference on Image Information Processing*, (ICIIP 2011)

[Patel (2018)] Nidhibahen Patel . *Dr. Selena (Jing) He,A Survey on Hand Gesture Recognition Techniques,Methods and Tools*, June 2018. 6.

[Shreyashi Narayann Sawant and Kumbhar ()] M S Shreyashi Narayann Sawant , Kumbhar . *Real time sign language recognition using pca, international conference on advanced control and computing technologies*, 2014.

[Mandeep Kaur Ahuja and Singh ()] *Static Vision Based Hand Gesture Recognition Using Principal Component Analysis, Innovation and Technology in Education (MITE)*, Amardeep Mandeep Kaur Ahuja , Singh . 2015. IEEE. (3rd Interational Conference on MOOCs)

[Rautaray ()] 'Vision Based Hand Gesture Recognition for Human Computer Interaction: A survey'. S S Rautaray , A . *Springer Transaction on Artificial Intelligence Review* 2012. p. .

[Xu (2017)] Pei Xu . *A Real-time Hand Gesture Recognition and Human-Computer Interaction System, arXiv.org*, 24 April. 2017. Cornell University