DOI: 10.21917/ijivp.2020.0328

HOG-BASED EMOTION RECOGNITION USING ONE-DIMENSIONAL CONVOLUTIONAL NEURAL NETWORK

J. Sujanaa and S. Palanivel
Department of Computer Science and Engineering, Annamalai University, India

Abstract

This work proposes an emotion detection approach using Histogram of Oriented Gradients algorithm. Emotion detection is a crucial area since the emotions are extremely person dependent and finding them is hard with various lightning and illumination changes. Most of the works in this field focus on predicting the emotion using the facial region. In the proposed work, emotion detection is done using the mouth region. The dataset is comprised of mouth images containing emotions such as happy, normal and surprised in the form of video frames. The mouth regions are detected using the Haar-Based Cascade classifier at 20 frames per second. The HOG features are then extracted to detect three emotions namely Happy, Normal and Surprised. These HOG features are then trained using One-Dimensional Convolutional Neural Network (1D-CNN). The experimental results show that the proposed system can identify the emotions which gave improved performance than the earlier works.

Keywords:

Convolutional Neural Networks, Emotion Recognition, Histogram of Oriented Gradients, Mouth Detection

1. INTRODUCTION

Facial expression recognition (FER) [1] have attained recent importance since identifying the internal thoughts of humans can be identified with the help of Artificial Intelligence (AI) capabilities. Emotion recognition has a significant role in identifying human thoughts with the help of machine learning capabilities. Emotion can be identified to study the behaviour or internal thoughts of the human. According to the psychologist, emotions are mainly for short time and mood is milder than emotion and it is a long-lasting one. The psychologist states that there are six types of emotions namely: happy, normal, sad, surprise, fear, anger and disgust [2]. The emotions can be expressed when there is no need for verbal communication [3]. The Facial Action Coding System (FACS) is used to identify the facial emotions based on the action units assigned to each facial component [4]. In text-based emotion recognition, the text documents are scanned for the emotion text within the documents and this domain is said to be natural language processing [5]. In speech-based emotion recognition, the utterances of a few words are detected and then classifies the emotion based on the presence of the particular sound signal [6]. The human face is the easiest way to suddenly identify an emotion-based the variation in the facial muscles. The human face is the easiest way to suddenly identify an emotion-based the variation in the facial muscles. There are several methods to identify facial emotions, namely methods based on geometric features. The landmarks points identified in the facial region and assigned for an emotion. Some of the methods focus on the area or facial patches to identify the emotion. Several techniques exist in machine learning methods like support vector machines (SVM) [7], K-nearest neighbours (KNN) [8], Decision trees, Random Forest etc. to identify FER. Emotion recognition is the art of mining the information available in the spatial-temporal region of the image. The appearance-based features like HOG [9], Local binary pattern (LBP) [10], Haar-based approaches, etc. have been used in FER for feature extraction process. The main aim of the work is to recognize the emotion using mouth region and HOG features using the deep learning methods instead of the traditional machine learning approaches. FER approaches to simplify the need for complex human inspection with the help of deep learning approaches. FER can be widely used in monitoring the emotions in feedback-based e-learning platforms, Music recommendation systems based on human emotions and mood etc. [11].

2. LITERATURE REVIEW

Several studies have been already made with FER approaches. Still, it remains a challenging task due to the complex human emotion patterns and identification with machine learning approaches. The facial expression recognition consists of four main steps namely: (a) identifying the region of expression (b) extracting the intrinsic values (features) from the region (c) Normalization phase to eliminate the wide changes of values in the dataset. (d) Classification techniques to identify emotions. In dataset collection, the facial region is identified and preprocessed. Then it is given to the feature extraction block. The features are then fed to any of the classification algorithms like SVM [12], Decision trees [13], Random forest [14], etc. and then the performance of the system is analyzed. Over the past years, CNN has given promising results in various classification techniques.

Joseph Juliana and Sharmila [15], researched and identified the three emotions happy, sad and angry using HOG and LBP features from facial components like nose, eyes and mouth and they trained the conventional neural network classifier using texture features and obtained an accuracy of 87.00% and 64.00% for HOG and LBP features respectively. The authors Santhosh and Sharma used a fusion of HOG and LBP features to identify the facial emotions which recognize the emotions with an accuracy of 96.20%. The author Swinkels et al. identified the emotions, where they used 19 key-points to extract HOG features and modelled with SVM classifier and obtained an accuracy of 89.78% [16]. In a video frame based emotion recognition system [17], a method based on LBP along with the Adaboost algorithm is used to read the Linear Binary Pattern (LBP) features and then fed to Gaussian Mixture Models (GMM) for emotion classification and this model gives maximum accuracy and minimum time consumption. The authors [7] Ragb and Asari, analyzed the new method of HOG by implementing the phase congruency technique for HOG features. The phase congruency is used to study the edges and corners of the images and the histogram of oriented phase gradients is combined and modelled as Histogram of Oriented Phase Gradients (HOPG) features and these are fed to SVM classifier to study the performance and they obtained an accuracy of 94.92%. The authors, [18] investigated that HOG algorithm for Handwritten Bengali Numeral Recognition. The HOG features are extracted and combined the manually acquired colour histogram features with it where the SVM classifier is used and obtained an accuracy of 98.04%. The HOG-PCA method gives them an accuracy of 65.80%. The fusion of HOG-LBP features gave an improved performance of 91.60% in hand gesture recognition methods [19]. Similarly, this kind of HOG-LBP features obtained for facial segmentation method with Artificial Neural Network (ANN) classifier gives an accuracy of 96.97% in [1]. Based on the literature examined, it is identified that HOG features can capture effective features than the other texture-based methods with varying bin size, orientation, pixels per_cell and cells_per_block. The HOG features can be utilized in the proposed work to extract highly discriminative feature vector to feed 1D_CNN. This method will aid the system to recognize the emotions.

In this work, the basic type of expressions like happy, normal and surprised are considered in the mouth regions to detect the emotion efficiently since the other expressions are very hard to find through mouth region. The mouth regions contain the most changing muscles like jaw upliftment while surprised, teeth visible or partially visible during Happy and no movements of the jaw and no visibility of teeth during Normal expression.

3. PROPOSED METHODOLOGY

The objective of this work is to detect the emotion using HOG features with 1D-CNN. The system consists of three main parts: dataset collection, HOG feature extraction, training and validation, testing. The proposed approaches use Scikit learn library [20] for pre-processing and extracting HOG features from the dataset, Keras [21] with TensorFlow [22] as backend library for all deep learning modelling. In this work, Haar based classifier is used to detect the mouth region in every video frame. The Fig.1 shows the overall framework of the proposed work. The input of the proposed system is in the form of video frames from the live webcam feed. The mouth region is detected using Haar classifier and pre-processed to extract the HOG features to feed to 1D-CNN training.

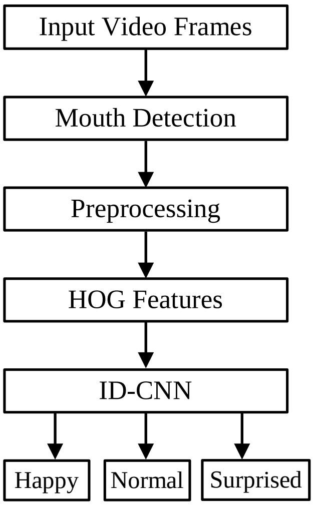


Fig.1. Framework of the proposed system for emotion recognition

In Fig.1 shows the mouth region is detected using the input video frames, then converted to grayscale images. The Scikit-learn library is applied for pre-processing where the RGB images are converted to grayscale images. The HOG algorithm is then applied to these images to capture the discriminative features. The performance of this proposed system is evaluated on the 1D-CNN classifier. The proposed system detects the three emotions namely happy, normal and surprised.

3.1 MOUTH DETECTION

The mouth region from the live webcam feed is detected using the pre-trained Haar cascade based OpenCV library [23]. This library consists of all the XML files for detecting eyes, nose and mouth region from the given input image or video frame. This cascade classifier is an algorithm developed by Paul Viola and Michael Jones using machine learning techniques [24].

3.2 PRE-PROCESSING

The detected mouth region is resized to 43X72 RGB image. During pre-processing "OpenCV-Contrib-python", machine learning library is used to convert the RGB[25] into single-channel gray-scale images[26] of the size 43×72×1 (No. of rows × no. of columns × no. of channels). Thus, the mouth-region images are converted to grayscale for faster processing of emotions. These images are fed to the feature extraction step. The Fig.2 shows the detected image and its corresponding grayscale image. The gradient of the image in X-direction and Y-direction is also shown Fig.2(c) and Fig.2(d) respectively.

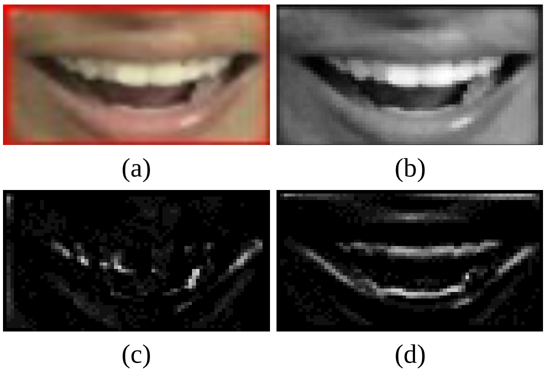


Fig.2. Histogram of Oriented Gradients (a) Detected mouth emotion using Haar Cascade Classifier, (b) Grayscale image (c) X-gradient image (d) Y-gradient image.

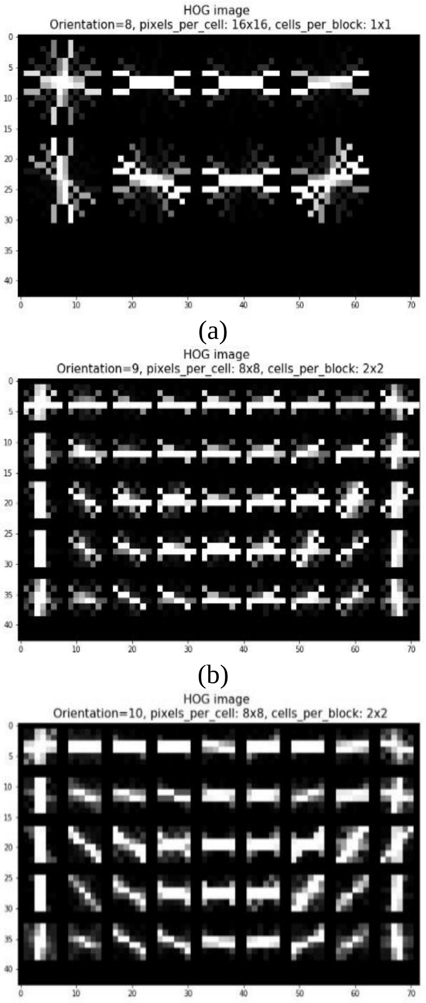
3.3 HOG FEATURE EXTRACTION

HOG is one of the popular features for extracting the texture-based features from the images. It is used to obtain the local features from the input images. HOG was initially used for human recognition developed by Dalal and Briggs [9]. It is the most robust texture features in terms of illumination and invariance. HOG is a powerful method to detect pedestrians and objects [18]. The workflow of the HOG algorithm is as follows:

- The entire input images are divided into cells and blocks, where the default cell size is 8×8 pixels and 2×2 cells in a block.
- The gradient magnitude and direction are computed for each pixel in the cell within the block.

- For each gradient magnitude, the corresponding bin is assigned based on the histogram gradients.
- For every block, normalization is done to eliminate the illumination changes.
- All the blocks are finally concatenated to form the histogram feature vector of ‘*n*’ dimension which is said to be the N-dimensional HOG features.

The Fig.3(a) shows the HOG image for the parameters, orientation = 8, pixels per cell = 16×16 and cells per block = 1×1 for input image. Similarly, Fig.3(b) shows the HOG image for the parameters, orientation = 9, pixels per cell = 8×8 and cells per block = 2×2 for input image, Fig.3(c) shows the HOG image for the parameters, orientation = 10, pixels per cell = 8×8 and cells per block = 2×2 for input image.



(a)

(b)

(c)

Fig.3. HOG image for the input happy emotion with different orientations

3.4 1D-CNN

CNN [27] belongs to the family of Deep Convolutional Neural Networks (DCNN) [28]-[30], where it is stacked with various convolutional layer and max-pooling layers sequentially. The input to CNN can be 1D, 2D or 3D array representation. 1D-CNN models are capable of deriving the input representation from the fixed-length array segments. Unlike the 2D CNN using matrix operations, these models use only single array operations. 1D CNNs have demonstrated superior performance on applications that have a limited labelled data and high signal variations acquired from different sources (i.e., patient ECG, civil, mechanical or aerospace structures, high-power circuitry, power engines or motors, etc.). In the convolution layer, the kernel or filter of size will slide over the input to perform multiplication or dot product operation to learn important features such as edges and outlines. When the array data to be one-dimensional, multiplication is performed element-wise and the result is summed up to form the output of this layer. In the sub-sampling layer, the kernel or filter of size will pass over the input from the convolution layer to perform dimensionality reduction which entails decreasing the matrix size by preserving essential information in the feature maps. In 1D-CNN [31] applies to data single dimension e.g. an array of features and the kernel moves along one direction. The structure of 1D-CNN is shown in Fig.4. These models take the input data in the form of a one-dimensional array and it is widely used in the analysis of time-series data, signal data processing and natural language processing (NLP). Fig.4. Shows an example of 1D-CNN architecture used where it is stacked with 4 convolution layers followed by 4 max-pooling layers with ‘*n*’ input feature vector length as input and with three output classes for recognizing three different emotions. The *n* is the number of HOG features obtained by varying its parameters.

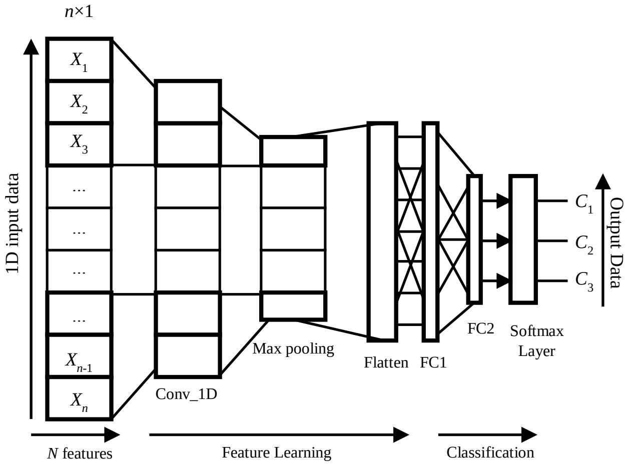


Fig.4. Structure of 1D-CNN architecture

The 1D-CNN is a model with input shape as (batch size, channel). The one-dimensional feature vectors are the input to this CNN model and the output has 3 classes. For every training input pattern, 1152 HOG features are extracted and so the training set consisting of 5376 samples will be converted into NumPy array shape as (5376, 1152, 1) and validation samples as (1344, 1152, 1) to feed to 1D CNN.

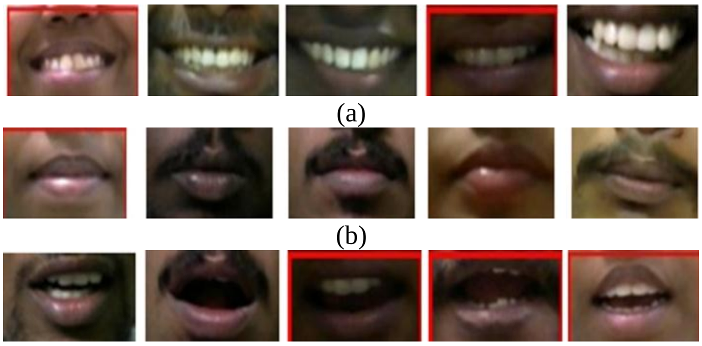
- Layer 1: In the first layer, 64 convolution filters with kernel size 3 are applied and the reduced convolved output will have 64 features arrays with size as 1150×64, followed by max-pooling with kernel size 3 is applied and the output of this layer will be 575×64.
- Layer 2: In the second layer, 64 convolution filters to produce a reduced feature map of size 573×64, followed by max-pooling with the same kernel size is applied and the size of the output feature map will be 286×64.

- Layer 3: In the third layer, 16 filters with kernel size 2 are applied to produce reduced feature map of size 285×32 and followed by a max-pooling layer with kernel size 2 is applied to get the reduced feature map of size 142×32.
- Layer 4: Similarly, the fourth layer contains 32 convolution filters and a max-pooling is applied having the same kernel size used in the previous layer. The final reduced feature map size will be 70×32. All the layers use stride with size 1.
- Flatten: Then the reduced feature vector is flattened to form 2240 values and fed as input to fully connected layers having varying neurons in each layer.
- Fully Connected and Output layer: The FC1 and FC2 have 100 and 50 neurons respectively and the last layer is the output layer in which 3 neurons are used with the Softmax activation function to classify the outputs into three classes namely happy, normal and surprised.

4. EXPERIMENTAL RESULTS

4.1 DATASET

The dataset is collected using a web camera with a resolution of 1280×720 in the laboratory environment. A total of 8400 mouth images are used from 20 subjects (10 Male, 10 Female). All the RGB images are converted to grayscale to fed to the CNN. The dataset is divided into training, validation and test sets. For training the architectures, 5376 mouth emotion images are used, 1344 mouth emotion images are used for validation and 1680 mouth emotion images are used for testing. Fig.5. Shows the samples of the dataset collected for happy, normal and surprising emotions.



(a)

(b)

(c)

Fig.5. Dataset (a) Happy (b) Normal (c) Surprised

4.2 FEATURE EXTRACTION

The feature extraction step is a primary step for finding out the most relevant features and omitting the unwanted information from the input images. The HOG feature with varying dimensions like 64, 96, 405, 1152, and 1280 are extracted and are fed to 1D-CNN for training. The Table.1 shows the number of HOG features obtained for varying the parameters like pixels per cell and cells per block and orientations.

Table.1. No. of HOG features obtained for varying parameters

No of pixels per cell	No of cells per block	Orientations	No. of HOG features
8×8	2×2	10	1280
8×8	2×2	9	1152
8×8	2×2	8	1024
8×8	1×1	9	405
8×8	1×1	8	360
16×16	2×2	10	120
16×16	2×2	9	108
16×16	2×2	8	96
16×16	1×1	9	72
16×16	1×1	8	64

4.3 TRAINING AND VALIDATION

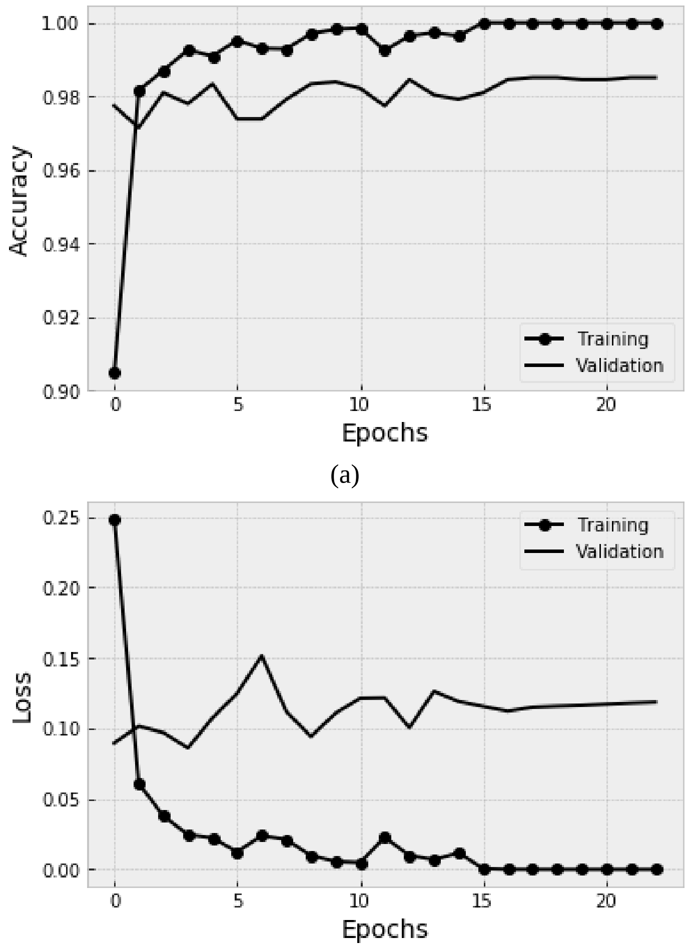
The features from the HOG descriptor is fed as input to the 1D-CNN model to detect emotion where the model learns the internal structure of data. The advantage of the use of 1D-CNNs for classification mainly prompts learning from the simple one-dimensional array data structure directly where there is no correlation within the array values.

The HOG descriptor is chosen since it can capture the feature effectively in entire human recognition. So, they can be used for emotion to capture the useful and important features from the dataset. The Table.2 shows the loss and accuracy for the varying HOG dimensions with the 1D-CNN model during training (Tr) and validation (Val). The HOG features with 9 orientation, 8x8 pixels per cell and 2×2 cells per block produced 1152 features, gives the maximum validation accuracy of 98.51%. For 1280 HOG features, the system produced a lesser accuracy of 98.21%. The system does not improve by increasing an additional number of HOG features and it produces efficient accuracy for 1152 HOG features obtained using 8×8 pixels per_cell and 2×2 cells in a block and having 9 bin orientation.

Table.2. Loss and Accuracy for training and validation data using the 1D-CNN model

No. of HOG features	Training Time (mm:ss)	Tr loss	Tr accuracy	Val loss	Val accuracy
1280	04:10	0.0550	0.9821	0.0767	0.9821
1152	**11:17**	**4.0125e^{-5}**	**0.1000**	**0.1150**	**0.9851**
1024	06:30	0.0035	0.9987	0.0838	0.9805
405	03:04	0.0200	0.9927	0.1801	0.9679
360	01:35	0.0629	97.86	0.1785	0.9600
120	01:22	0.0375	98.72	0.1243	0.9744
108	00:42	0.0970	97.05	0.1229	0.9673
96	00:41	0.0880	96.98	0.1212	0.9667
72	01:09	0.0355	98.75	0.1801	0.9506
64	00:58	0.0824	97.10	0.2188	0.9464

In Fig.6(a) shows the accuracy graph for training and validation accuracy with 1152 HOG features and the model stops at 17 epochs in a CPU with 1.2GHz and 8 GB RAM and Fig.6(b) having its corresponding loss curve. The total training time for the system to learn the input parameters is 11 min and 17 sec for 1152 HOG features.



(a)

(b)

Fig.6. Training and validation data using 1D-CNN for 1152 HOG features. (a) Accuracy plot (b) Loss plot

4.4 TESTING

The testing is an important part of any application to analyze the performance of the trained models. Therefore, the 1680 testing samples are given to the trained model using 1D-CNN architecture. The confusion matrix is the most efficient way to identify the True Positives (*TP*), True Negatives (*TN*), False Positives (*FP*), False Negatives (*FN*) and Accuracy (*ACC*) of a classifier and it is used for classification problems having binary or multi classes associated with the output. The true positive are the number of correctly classified samples as the labelled class. The False positives are the number of samples mistakenly classified as the labelled class. True negative is the samples correctly identified as labelled emotion in other classes. False negatives are the incorrectly classified emotion to other classes. The accuracy (*ACC*) is defined as the total number of corrected identified emotions to the total number of samples and it is given by,

$$ACC = [(TP+TN)]/[(TP+TN+FP+FN)] \quad (1)$$

The precision is the ratio of *TP* with *TP* and *FP* and it is calculated as,

$$P = TP/(TP+ FP) \quad (2)$$

The recall is the ratio of TP with TP and FN and it is calculated as,

$$R = TP/(TP+ FN) \quad (3)$$

The precision and recall are combined to form the measure called F1-score, which is the harmonic mean of *P* and *R*.

$$F = 2*[(precision*recall)/(precision+recall)] \quad (4)$$

In confusion matrix [32], true positive are the total number of emotion frames which are correctly identified as the respective classes, true negatives are the number of emotion frames which are correctly identified as other classes. The confusion matrix for every 10 consecutive testing images is shown in Fig.7. The model classifies 89.60% happy samples is correctly classified as happy classes in testing, 91.96% testing as a normal class and 89.50% testing as a surprised class respectively. In total 90.23% is correctly classified and 9.77% is in-correctly classified.

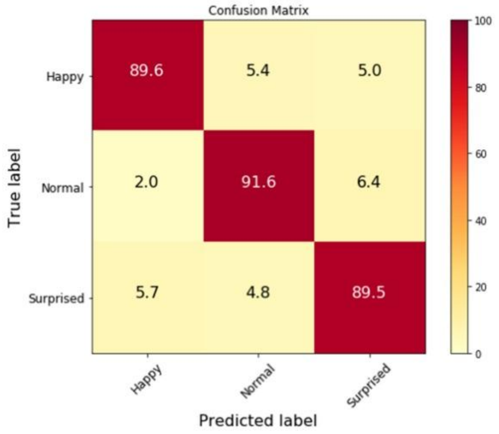


Fig.7. Confusion Matrix for testing data

5. CONCLUSION

In this work, we have made a comprehensive study of different scales of HOG features to identify the emotions is made. This work proposed an emotion recognition system for recognizing the three emotions normal, happy and surprised using 1D architecture. The mouth video frames are extracted in an unconstrained laboratory environment using a web camera. The HOG can efficiently capture the important highlights from the emotion images. This work can further be evaluated on other emotion datasets like CK+, MMI, FER2013 etc. to study its performance. The proposed system recognizes the three emotions with an accuracy of 90.23%.

REFERENCES

[1] B. Islam, F. Mahmud, A. Hossain, M.S. Mia and P.B. Goala, "Human Facial Expression Recognition System using Artificial Neural Network Classification of Gabor Feature based Facial Expression Information", *Proceedings of 4th International Conference on Electrical Engineering*

Information Communication Technology, pp. 364-368, 2019.

[2] P. Ekman, "*Basic Emotions*", Handbook of Cognition and Emotion, 2005.

[3] A. Bhavan, P. Chauhan and R.R. Shah, "Bagged Support Vector Machines for Emotion Recognition from Speech", *Knowledge-Based Systems*, Vol. 184, pp. 104886-104896, 2019.

[4] V.M. Alvarez, R. Velazquez, S. Gutierrez and J. Enriquez-Zarate, "A Method for Facial Emotion Recognition Based on Interest Points", *Proceedings of IEEE International Conference on Research in Intelligent and Computing in Engineering*, pp. 1113-1122, 2018.

[5] E. Batbaatar, M. Li and K.H. Ryu, "Semantic-Emotion Neural Network for Emotion Recognition from Text", *IEEE Access*, Vol. 7, pp. 111866-111878, 2019.

[6] J. Chen, Z. Chen, Z. Chi and H. Fu, "Facial Expression Recognition in Video with Multiple Feature Fusion", *IEEE Transactions on Affective Computing*, Vol. 9, No. 1, pp. 38-50, 2018.

[7] H.K. Ragb and V.K. Asari, "Histogram of Oriented Phase and Gradient (HOPG) Descriptor for Improved Pedestrian Detection *Proceedings of IEEE International Conference on Electronic Imaging Science and Technology*, pp. 1-6, 2016.

[8] D. Reney and N. Tripathi, "An Efficient Method to Face and Emotion Detection", *Proceedings of IEEE International Conference on Communication Systems and Networking Technology*, pp. 493-497, 2015.

[9] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection", *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 886-893, 2005.

[10] T. Ahonen, A. Hadid and M. Pietikäinen, "Face Recognition with Local Binary Patterns", *Proceedings of IEEE International Conference on Computer Vision*, pp. 23-29, 2004.

[11] F. Abdat, C. Maaoui and A. Pruski, "Human-Computer Interaction using Emotion Recognition from Facial Expression", *Proceedings of IEEE International Conference on Computer Modelling and Simulation*, pp. 196-201, 2011.

[12] L. Zhu, L. Chen, D. Zhao, J. Zhou and W. Zhang, "Emotion Recognition from Chinese Speech for Smart Affective Services using a Combination of SVM and DBN", *Sensors*, Vol. 17, No. 7, pp. 2-14, 2017.

[13] L. Sun, B. Zou, S. Fu, J. Chen and F. Wang, "Speech Emotion Recognition based on DNN-Decision Tree SVM Model", *Speech Communications*, Vol. 115, pp. 29-37, 2019.

[14] M. Arora, M. Kumar and N. Kumar, "Facial Emotion Recognition System Based on PCA and Gradient Features", *National Academy Science Letters*, Vol. 41, No. 6, pp. 365-368, 2018.

[15] J.K. Josephine Julina and T.S. Sharmila, "Facial Emotion Recognition in Videos using HOG and LBP", *Proceedings of IEEE International Conference on Electronics, Information, Communication and Technology*, pp. 56–60, 2019.

[16] W. Swinkels, L. Claesen, F. Xiao and H. Shen, "Real-Time SVM-Based Emotion Recognition Algorithm", *Proceedings of IEEE International Conference on Image and Signal Processing, Biomedical Engineering and Informatics*, pp. 1-6, 2018.

[17] H. Kaya, F. Gurpinar and A.A. Salah, "Video-Based Emotion Recognition in the Wild using Deep Transfer Learning and Score Fusion", *Image and Vision Computing*, Vol. 65, pp. 66-75, 2017.

[18] S. Wang, L. Duan, C. Zhang, L. Chen, G. Cheng and J. Yu, "Fast Pedestrian Detection based on Object Proposals and HOG", *Proceedings of IEEE International Conference on Neural Networks*, pp. 3972-3977, 2016.

[19] H. Lahiani and M. Neji, "Hand Gesture Recognition Method based on HOG-LBP Features for Mobile Devices", *Procedia Computer Science*, Vol. 126, pp. 254-263, 2018.

[20] F. Pedregosa, "Scikit-Learn: Machine Learning in Python", *Journal of Machine Learning Research*, Vol. 12, pp. 2825-2830, 2011.

[21] Keras, Available at: https://keras.io/, Accessed at 2020.

[22] Martin Abadi et. al., "Large-Scale Learning on Heterogeneous Distributed Systems", *Proceedings of IEEE International Conference on Distributed, Parallel, and Cluster Computing*, pp. 1-19, 2016.

[23] The OpenCV Reference Manual, Available at: http://www.cs.unc.edu/Research/stc/FAQs/OpenCV/OpenCVReferenceManual.pdf, Accessed at 2010.

[24] M.J. Paul Viola, "Rapid Object Detection using a Boosted Cascade of Simple Features", *Pattern Recognition*, Vol. 39, No. 3, pp. 1-9, 2001.

[25] RGB Color Model-Wikipedia, Available at: https://en.wikipedia.org/wiki/RGB_color_model, Accessed at 2020.

[26] Grayscale-Wikipedia, Available at: https://en.wikipedia.org/wiki/Grayscale, Accessed at 2020.

[27] L. Yann and B. Yoshua, "*Convolutional Networks for Images, Speech, and Time-Series*", The Handbook of Brain Theory and Neural Networks, Vol. 4, pp. 2571-2575, 1998.

[28] A. Krizhevsky, I. Sutskever and G.E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", *Proceedings of IEEE International Conference on Advances in Neural Information Processing Systems*, pp. 1178-1185, 2012.

[29] A. Krizhevsky, I. Sutskever and G.E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", *Communications of the ACM*, Vol. 60, No. 6, pp. 1-18, 2017.

[30] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions", *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1800-1807, 2017.

[31] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj and D.J. Inman, "1D Convolutional Neural Networks and Applications: A Survey", *Mechanical Systems and Signal Processing*, Vol. 151, pp. 1-20, 2019.

[32] Kavita Ganesan, "Text Mining, Analytics and More: Computing Precision and Recall for Multi-Class Classification Problems", Available: https://kavita-ganesan.com/how-to-compute-precision-and-recall-for-a-multi-class-classification-problem/#.XddU1TJKhhE, Accessed at 2019.