INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

IJARIIT

ISSN: 2454-132X
Impact factor: 4.295
(Volume 5, Issue 4)
Available online at: www.ijariit.com

A facial-expression recognition model using deep learning

Shreyas Limaye
slimaye24@gmail.com
MIT College of Engineering, Pune, Maharashtra

Dr. Virendra. V. Shete
virendra.shete@mitcoe.edu.in
MIT College of Engineering, Pune, Maharashtra

ABSTRACT

Deep neural networks have been recently putting a breakthrough in pattern recognition, machine learning and artificial intelligence. This paper emphasizes a study based on deep learning framework contributing to the field of expression recognition. The proposed model involves a technique using deep for human facial expression recognition. Images are first preprocessed with normalization manipulation to remove illumination and facilitate enhancement using hat-filtering. At that point, a weighted, focus symmetric nearby paired example (CS-LBP) is connected to each face hinder by piece. The CS-LBP pieces are connected to form an element vector of the face picture. The deep network is trained using the layer-wise strategy. We use the CIFAR-10 dataset is used for training and testing. A database of real images is used for testing the algorithms. GUI has been created which compares trained and tested dataset and specifies the type of expression in the command window.

Keywords*— Deep learning, Expression recognition, CS-LBP*

1. INTRODUCTION

Facial expression recognition, sets its benchmark for automatic facial expression recognition using various features. It has many applications in Human Computer Interaction (HCI), artificial intelligence, security monitoring, social entertainment, etc.

Even though there are many ways to investigate the recognition of humans' emotional behavior, ranging from geometric features, posture of the body, speed and voice pitch, in this paper we focus only on one area i.e. visual expression recognition of images. Expressions play an important role people's everyday life. So, it is very important to monitor these changes as they contain information about different types of emotions which will assist in understanding behaviors. There are six basic universal expressions viz., happiness, surprise, fear, sadness, anger and disgust. These emotions are associated with specific facial expressions.

One of the main challenges in emotion recognition as in most computer vision tasks is to deal with the complexity of real-world scenarios. This includes large variations of illumination, appearance of subjects and sensor properties, among many other parameters. A variety of image processing methods have been applied to face recognition but the more popular methods include Principal Component Analysis (PCA), Support Vector Machines (SVM) and Artificial Neural Networks (ANN). By using ANN, the differences of feature data between faces are clarified and recognized. The demand for large data sets is increased with the inception of deep learning models.

In this project, we develop a pre-processing and enhancement technique for facial attribute recognition applications. The proposed method uses the value of monochromatic images and extract different facial features. The extracted information is fed into input layer of a deep neural network model for face identification. On the other hand, the color images are used by the recognition algorithm to eliminate non-skin colored background and reduce further processing time.

2. RELATED WORKS

Muhammad [1] proposed a facial-expression recognition model to facilitate the healthcare service in a smart city. Bandlet transform has been applied to face image block by block and features were extracted using CS-LBP. GMM and SVMs were used for classifying the most dominant features.

Mehrabian [2] indicated that spoken words (7%), voice (38%) and facial expressions (55%) are accustomed to convey messages by humans. Extraction of face expression is the core step of derivation of expression recognition. The two necessary forms of approaches to extract features vectors are: appearance-based strategies and geometric based strategies. My motivation is that face images can be seen as a composition of smaller patterns on that LBP can be applied.

The paper [3] described a facial expression recognition system that extracted the users' attributes using facial actions and head motion features. (LBP) was deployed for texture analysis and then extended for other applications. The recognition stage grey scale images using 2D-DCT for future processing.

Tivatansakul et al [4] developed a healthcare system, which focused on the expression recognition to recognize and tackle with negative emotional health in daily life. They used the facial expression and vocal data to find out the users' emotions using a directional ternary pattern (DTP). If a negative emotion was detected, a relaxation activity to the user was suggested.

Tian et al. [5] posed a Neural Network primarily based approach to greet facial action units in image sequences. Cohen et al. [6] projected a multi-level HMM classifier that performs expression classification on a video phase, but also phases out a protracted video sequence to the various expressions segments. However, HMMs cannot handle dependencies in observation.

Feature extraction of facial expressions is the main problem on facial expression recognition tasks [7]. The extracted features are vital for classification performance. The aforementioned hand designed methods depend on manual operations with labelled data for feature extraction. Deep learning has a plethora of features that can be directly learned from original unsupervised data.

The structure of paper is as follows: Section 3 illustrates the proposed architectural flow for training, testing and classification, followed by CS-LBP algorithm analysis for feature extraction with each and every step involved in the extraction process. The paper concludes with partial experimental results showing a GUI for detection and recognition of proposed expressions of the called facial function.

3. PROPOSED ARCHITECTURE

Figure 1 demonstrates a square outline of the proposed outward appearance acknowledgment framework. Input images have been taken from MDBI database trained using CIFAR-10 dataset.

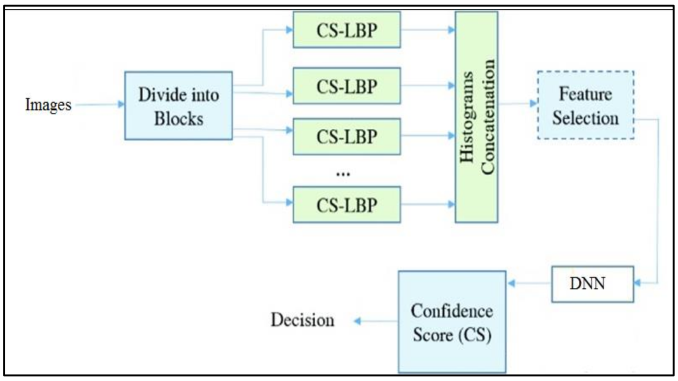

Fig. 1: Block diagram of proposed system

Our facial expression recognition analysis can be done in three main steps:
(a) Face acquisition;
(b) Facial feature extraction, and;
(c) Expression recognition.
The proposed system is implemented based on the following steps:

3.1 Image acquisition and pre-processing

The images to be trained have been procured from a static MDBI database with 10 facial expressions. Here pre-processing has been done for de-noising of the input images for better performance. For this, each image has been converted to grayscale by using 3D colour image arrays. This conversion compresses size of the image. Training ANN with monochromatic images decreases the no. of computations and memory requirements; thereby reducing the training time and increasing the speed in which a trained network can identify the image

3.2 Feature extraction

Extracting the best features is the prime focus of any successful facial expression recognition system. Effectiveness of the facial image representation could provide a boost for robust recognition process. The no. of features depends on the no. of blocks in the CS-LBP graph. If the number of features is high, it takes more time to take decision.

3.3 Classification and facial expression recognition

We start by training a deep neural network using images of over identities. Our face recognition model is also computationally inexpensive. Running all models combined takes less time compared to the standalone model and we achieve better results. Even though the network architecture is not the same for each task, all networks are trained for the task of facial recognition using the image set.

4. FACIAL EXPRESSION WITH DCNN

4.1 Normalization

The images in the database differ in a variety of parameters which can affect directly on accuracy and performance of the facial recognition. For every individual, rotation, brightness and illumination changes even for the same person's images. To tackle with this problem, normalization techniques such as detecting, de-noising and some other preprocessing such as top hat and bottom hat filtering along with image enhancement has been performed. The image brightness and contrast variations complicate the problem. That's the reason monochromatic images have been used, leaving the luminance data to comprise this image.

4.2 Image cropping

The original face images may have unnecessary information in the background and could distort the output. Cropping region is done to remove the redundant facial parts that do not contribute in the given expression.

4.3 Down sampling

Down sampling is done to ensure the same location all the face components in every face image. Down sampling helps DNN to learn which regions are related to each specific expression and also enables the training to be performed on the GPU more effectively.

4.4 CS-LBP pattern generation

The next step is to divide each into blocks. The CS-LBP is applied to each image pixel by pixel. After training, we see a 'traininglabels' dataset generated in our training source folder. Center symmetric pixels are subjected to comparative analysis based on their grayscale intensities, as illustrated in figure 1.

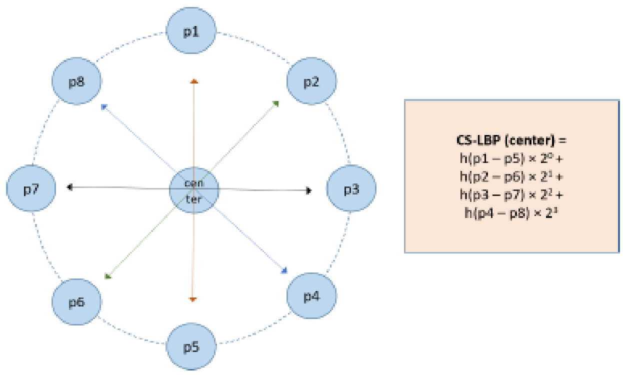

Fig. 2: Calculation of CS-LBP [1]

The calculation of the CS-LBP is expressed by equation (1).

$$CSLBP_{P,R} = \sum_{j=0}^{(P/2)-1} 2^j q(P_j - P_{j+(P/2)})$$

$$q(x) = \begin{cases} 1, x \geq T \\ 0, otherwise \end{cases} \qquad (1)$$

In the above equation, pj is the gray-scale intensity of the pixel (p). P and R are the numbers of pixels in a circular neighborhood, where the radius of the circle is R. In our work, we chose P = 8 and R = 1.

To maintain the spatial information in the CS-LBP histogram, it is calculated block-by-block [8]. The histogram of each block has been appended a weight. The weight of each block is calculated by the information entropy of the block, as follows:

$$u_m(h) = \sum_{(i,j)\in Block_m} R\left(CSLBP_{P,R}(i,j), h\right), h \in [0, L]$$

$$R(a,b) = \begin{cases} 1, & a = b \\ 0, & otherwise \end{cases}$$

$u_m(h)$ is the probability in the m-th block which (total bin is L+1) appears for a pixel. Following which, the entropy is calculated as follows.

$$E_m = \sum_{h=1}^{L} u_m(h) ln u_m(h)$$

The weight of the m-th block (where the total number of blocks is n) is calculated as follows:

$$w_m = \frac{E_m}{\sum_{m=1}^{n} E_m}$$

The weighted CS-LBP histograms from the blocks are concatenated generating a feature matrix of given image input. The number of features in the feature vector depends on the number of blocks in the CS-LBP histogram. Higher the features selected, more is the time it takes for reaching a decision.

5. FLOWCHART

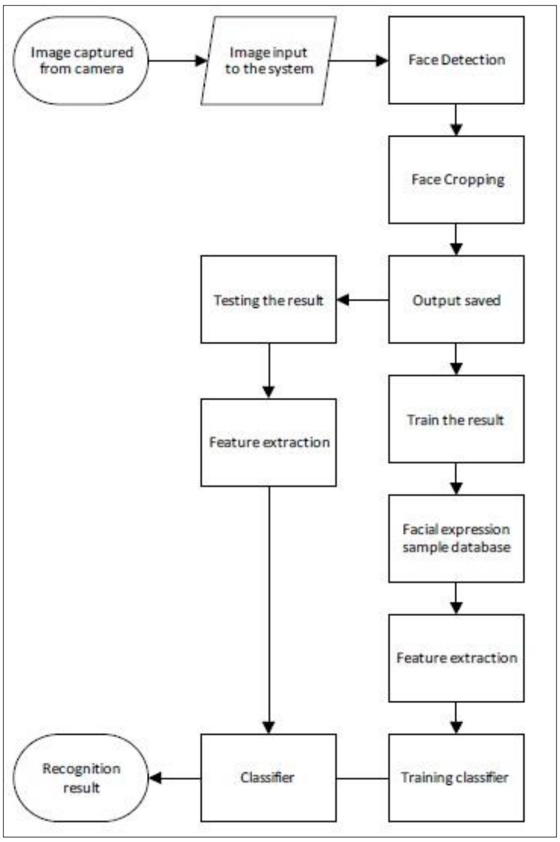

Fig. 3: Flow chart

6. PERFORMANCE METRIC

To validate any system it is necessary to check its accuracy and error rate. Here confusion matrix is used to calculate the accuracy. In machine learning and specifically the problem of statistical classification, a confusion matrix, also known as error matrix. Each column of the matrix represents the instances in a predicted class while each row represents the instances in an actual class or vice versa. In predictive analysis, a table of confusion is a table with two rows and two columns that reports the number of false positive, false negatives, true positive and true negative.

7. RESULTS

In our project, we have used CIFAR-10 dataset for training and testing purpose. Foll.is the CS-LBP graphical representation of a trained image.

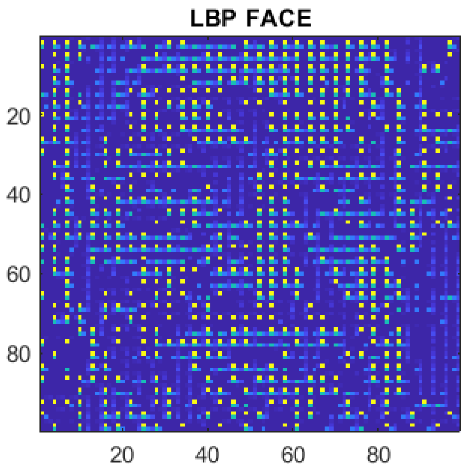

Fig. 4: CS-LBP representation of trained image generated in MATLAB

The table below shows the results of the detected/predicted images compared using 'FrontalFaceLBP' function with the trained images. Instead of recognizing only ten expressions, we can also recognize the neutral expression, which results in a classifer that recognizes ten expressions. The input image is processed by the 32 learned kernels and generates 32 output maps. Depending on the expression count recognized and images detected using confusion matrix, the accuracy was found to be 91.25%.

Table 1: Confusion matrix of the real-time predicted/tested image set

300	**Predicted Values**				
Actual Values	I/O	Happy	Impressed	Confused	
	Happy	77	19	4	100
	Impressed	1	81	18	100
	Confused	22	0	78	100
	Total	100	100	100	

Actual Input Images \ Predicted (Tested) images	Angry	Smile	Tired	Sad	Happy	Impressed	Confused	Fear	Insecure	Arrogant	Neutral
Angry	75	0	0	0	0	0	0	0	0	0	0
Smile	0	78	0	0	0	0	0	0	0	0	0
Tired	0	0	100	0	0	11	0	0	0	0	0
Sad	0	0	0	98	0	7	0	0	11	0	0
Happy	0	0	0	0	77	0	0	0	0	0	0
Impressed	0	0	0	0	0	81	0	0	0	0	0
Confused	0	0	0	0	0	0	68	0	0	0	0
Fear	0	0	0	0	0	0	0	31	0	1	0
Insecure	0	0	0	0	0	0	0	18	89	88	0
Arrogant	0	0	0	0	0	0	0	0	0	0	0
Neutral	35	0	0	2	23	1	32	51	0	11	100

Fig. 5: Confusion matrix of facial expression recognition of training set

7.1 Calculations for Accuracy

(a) True Positive (Happy) = 77
(b) True Positive (Impressed) = 81
(c) True Positive (Confused) = 68

Accuracy for the tested and predicted results using confusion matrix is calculated as follows:

$$\frac{\text{True Positve values for Expressions}}{\text{Total}}$$

$$= \frac{77+81+78}{300} = 78.66\%$$

7.2 Creation of GUI

A GUI will be created with respective databases showing human facial attributes and expressions showing the accurate expression classified using DNN as shown in figure 4.

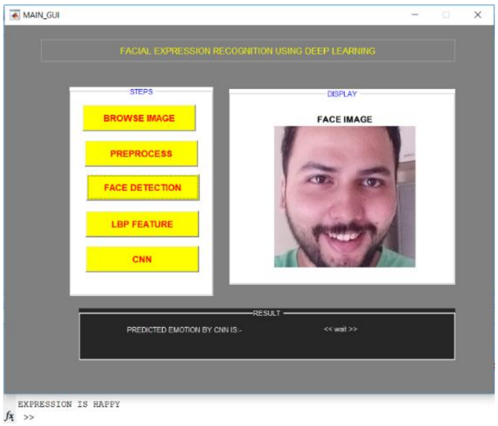

Fig. 6: Results in MATLAB GUI

8. CONCLUSION

In this paper, a brief description of facial expression recognition with DNN has been presented. The new learning technologies (especially DCNN) help to achieve better accuracy and performance. In future, we will broaden the work here by fusing electronic medicinal records into the proposed framework for better social insurance.

9. REFERENCES

[1] Ghulam Muhammad, "A Facial-Expression Monitoring System for Improved Healthcare in Smart Cities" , 2016 IEEE.
[2] Mehrabian, "Communication without words", Vol. 2, 1968.
[3] P.Ekman, "FACS: A Technique for the Measurement of Facial Movement", 1978.
[4] T. Achalakul, and M. Ohkura, ''Healthcare system focusing on emotional aspect using augmented reality: Emotion detection by facial expression, IEICE, vol. 114, no. 68, 2014.
[5] Tian et al.., "Approach on random weighted deep neural learning model for electricity customer classification," IEEE Wireless Communication magazine, vol. 22, no. 6, pp.67-75, Dec. 2015.
[6] Cohen, N. Sebe, A. Garg, L. Chen, T.S. Huang, "Facial expression recognition from video sequences", Computer Vision and Image Understanding 91, 2003.
[7] Xiaoming Zhao, "Facial Expression Recognition via Deep Learning", IETE Technical Review,2015.
[8] Castro F M, Marín-Jiménez M J, Guil N., "Multimodal features fusion for gait, gender and shoes recognition" Machine Vision and Applications, 2016:1-16

