JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Optimization Accuracy of Facial Emotion Recognition using Convolution Deep Neural Network

[1]Uday Shankar Dikshit, [2]Prof. Divya Jain
[1]M. Tech. Scholar, [2]Professor
Department of Electronics and Communication Engineering
TIT, Bhopal, India

Abstract: Facial Emotion Recognition is in pattern these days in view of its wide application zone. With the advancement in the field of profound learning, Facial feeling acknowledgment has become a significant subject in the business. There is critical examination on it since 1960; still, it is a difficult errand for the analysts in the business. Feeling acknowledgment is utilized to perceive the feelings of an individual. Different strategies have been created to recognize the feeling of an individual from still pictures and live recordings. At the same time, the basic principle of MLP Grasp the full connection layer and classification layer, and use Python's theano library to achieve. The construction and training of CNN model based on face recognition are studied. To simplify the CNN model, the convolution and sampling layers are combined into a single layer. Based on the already trained network, greatly improve the image recognition rate.

***Index Terms* – Facial Expressions, CNN, Trained Network**

I. INTRODUCTION

Facial expressions are the vital identifiers for human feelings, because it corresponds to the emotions. Most of the times (roughly in 55% cases) [1], the facial expression is a nonverbal way of emotional expression, and it can be considered as concrete evidence to uncover whether an individual is speaking the truth or not [2]. The current approaches primarily focus on facial investigation keeping background intact and hence built up a lot of unnecessary and misleading features that confuse CNN training process. The current manuscript focuses on five essential facial expression classes reported, which are displeasure/anger, sad/unhappy, smiling/happy, feared, and surprised/astonished [3]. The FERC algorithm presented in this manuscript aims for expressional examination and to characterize the given image into these five essential emotion classes. Reported techniques on facial expression detection can be described as two major approaches. The first one is distinguishing expressions [4] that are identified with an explicit classifier, and the second one is making characterization dependent on the extracted facial highlights [5]. In the facial action coding system (FACS) [6], action units are used as expression markers. These AUs were discriminable by facial muscle changes.

Outward appearances are utilized to accommodate the passionate state. As planned to give us an unmistakable image of what an individual resembles when the person has a passionate state in the external world. This see is sincerely and all the more firmly identified with an individual's inward state as feelings are the arrangement of a specific perspective, and watchers can undoubtedly survey an individual's otherworldly condition to help him or offer a response to his outward appearance[1]. Medical caretakers search for something like this in psychological well-being care offices to help Doctors in building up an accessible therapy plan for patients. So, it is vital to comprehend the Comprehension of feelings, the purposes for specific feelings, and various feelings in decisions, medical clinics, sentiment, schools, and guiding centers. Facial enthusiastic acknowledgment frameworks are utilized to distinguish outward appearances, which can be quite possibly the most perplexing undertakings or PC vision issues. Since it is hard to prepare a particularly model, the model ought to perceive the qualities that separate one feeling from resentment like Happy, Angry, Sad, Calm, Surprise, and so forth [2].The main purpose of human emotion recognition is to identify a user's emotions on the basis of certain data provided. Emotions are a response that stays for seconds. Emotional state refers to a person's current state (arousal, mood, or personality) regardless. The declared objectives can be realized by the many different divisions made in the model identification field.

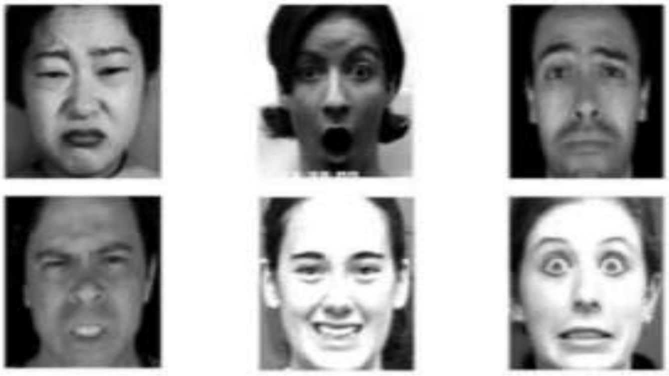


Fig.1: Human Facial Emotion

This method looks at different phases of taxonomy creation, such as data acquisition and attributes extraction, labeled data, and the creation of training sets with taxonomic practice. In our work, we believe that using a computer video camera as a standard input device will lead to emotional recognition based on multi-model input show in Fig.2. This method provides detailed facts and statistics about the identification method because different information sources can provide important related details by eliminating potential errors for each input. The purpose of this study is to introduce a facial-based approach to perceive a learner's understanding of the whole process of distance learning. This study proposes a model for the study of sensory learning, which consists of three phases: Feature detection, Feature extraction and classification of emotion show in Fig.1.

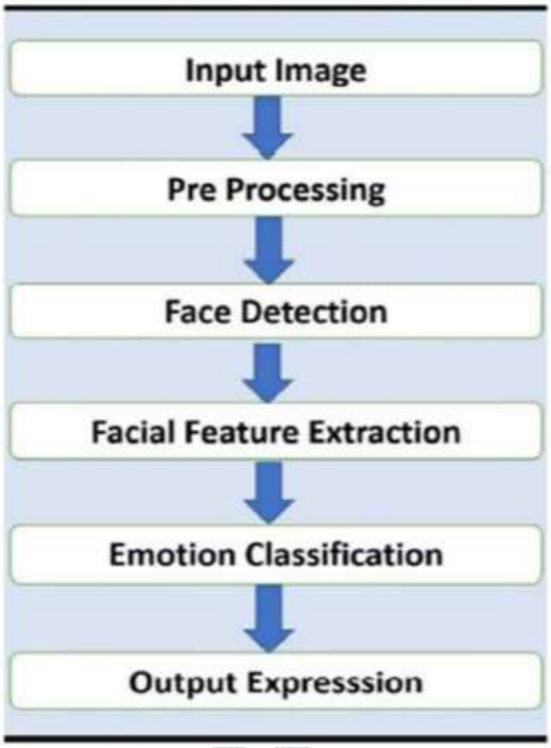


Fig.2: Classification Process

II. CONVOLUTION NEURAL NETWORK

Convolutional neural networks are presently among the best flamboyant algorithms for deep learning with image data. They have been successfully applied in computer vision tasks, and the robustness in object recognition localization in variant images is proven by the results. Thorough research on automatic analysis of expression was publicized recently [8], [9], [10]. These publications have established a set of standard algorithmic pipelines for FER. Though, their concentration is on classic methods, and deep learning has seldom been reviewed. In the recent time, FER has been examined according to deep learning in [11], but this is a very short review without introductions on FER datasets and technical details on deep FER. Hence, we carry out in this study, a systematized study on deep learning for FER duty that depend on videos (image sequences) and non-moving image. Aiming to provide new researchers is on this field a general description of the systematized architecture and high skills of deep FER. It consists of input and output layer.

Neural network can be divided into two kinds, biological neural network is one of them, and artificial neural network is another kind. Here mainly introduces artificial neural network. An artificial neural network is a data model that processes information and is similar in structure to the synaptic connections in the brain. Neural network is composed of many neurons; the output of the previous neuron can be used as the input of the latter neuron.

This unit is also called Logistic regression model. When many neurons are linked together, and when they were layered, the structure can now be called a neural network model. Fig. 3 shows a neural network with hidden layers.

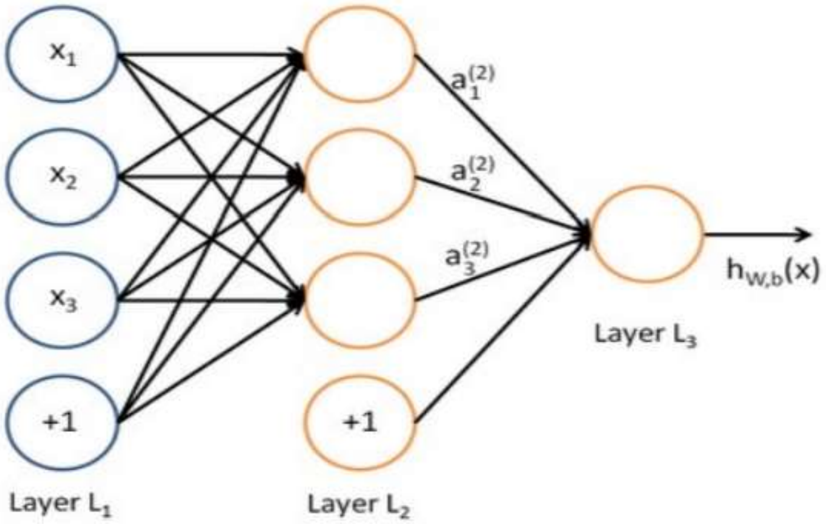


Fig. 3: Neural Networks

Convolutional Layer:- Convolutional layer has always been the first layer of CNN, it extracts features, and it consists of several feature maps. Each neuron in a feature map is connected to a small region, called the local receptive field, through a set of shared weights and a single shared bias. The main two advantages and reasons why Convolutional layer is preferable over fully connected layer are; firstly, parameter sharing; where all neurons share equal weights and bias in a feature map, which causes fast training as a result of a high reduction in the number of parameters, and will eventually help in building deep networks [11].

Pooling Layer:- After convolutional layer, the pooling layer is usually implemented with the aim of reducing the spatial resolution of the feature maps, speeding the computation and extracting prominent features. Max-pooling is the most used technique for pooling layer, where each pooling unit is equal to the largest element in a receptive field.

Fully Connected (FC) Layer:- The fully connected layer input must be a vector, so we have to first flatten the output features from the two layers (convolutional and pooling). Then, we may pass them to the output layer, where a Soft-Max classifier or sigmoid are used for predicting the input class label. Besides these networks, there are also much common derived architecture. In [12], region-based CNN (RCNN) [19] was utilized for learning features for FER. In [20], Faster R-CNN was employed to signify facial expressions through generation of proposals of a high class region. In addition, 3D CNN was introduced to acquire motion information encoded in various close structures for recognizing actions through 3D convolutions. The well-designed C3D was proposed which utilizes 3D convolutions on large-scale supervised training datasets to learn patio-temporal features. Several related researches have used this network for FER.

III. RESEARCH METHODOLOGY

At present, the typical architecture of neural network is divided into the following categories: LeNet5, AlexNet, ZF Net, GooLeNet, and VGGNet, the following will LeNet5 architecture for a detailed analysis. LeNet5 is a CNN classic structure that existed long ago, and it is mainly used in the recognition of handwritten fonts. It contains a total of seven layers of structure, except for the input layer, each of the other has training parameters, and each layer contains a plurality of Feature Maps, we can extract the input features through a convolution kernel. And each feature contains multiple neurons. The picture below shows the architecture of LeNet5:

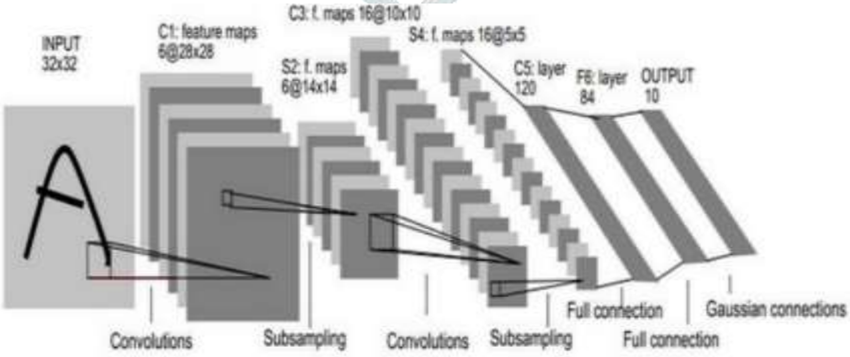


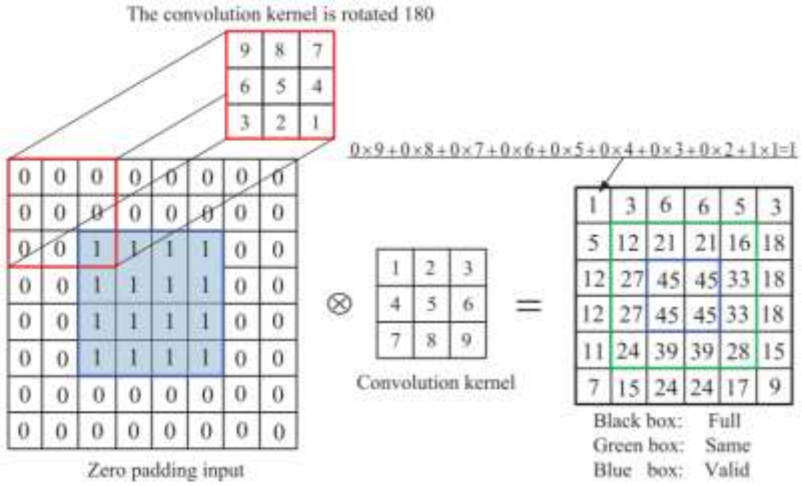
Fig. 4: LeNet5 structure diagram

As shown in Figure 4, a size of 32*32 images through the input layer into the network structure. The layer in the input layer is a convolution layer, which is represented by C1. The number of convolution kernels is 6 and the size is 5*5. After this layer processing, the number of neurons is 28*28*6, trainable parameters are (5*5+1)*6. The next layer of the C1 layer is a down sampled layer, shown in the figure, whose input is the output of the layer convolutional layer, 28*28 in size, 2*2 in the spatial neighborhood of the sample, and the way it is sampled Is to add 4 numbers, multiply them by a trainable parameter, and then add a trainable offset to output the result through the sigmoid function. The number of neurons in layer S2 is 14*14*6. After passing through the S2-layer sampling tube, the size of each feature plot it gets is a quarter of the output from its previous convolution layer. The layer after layer S2 is still a convolutional layer, with a total of 16 convolution kernels, and the size of each convolution kernel is the same as that of C1.

This layer is called the C3 layer in the above figure. The size of the output feature layer in this layer is 10*10. The 6 features in the S2 layer are connected with all the features in the C3 layer. The features obtained in this layer the figure is a different combination of the output features of the previous layer. The S4 layer is the same as the S2 layer, and its sampling type is 16. So far, the network structure has reduced the number of neurons to 400. The next layer of C5 is still a convolutional layer, which is fully connected with the previous layer, the size of its convolution kernel is still 5*5, this time C5 layer image processing, the image size becomes 5-5+1=1, which means that only one neuron output, in this layer contains a total of 120 convolution kernel, so the final output of neurons is 120. The last layer of F6, this layer is a fully connected layer, by calculating the input vector and the weight vector between the dot product, plus a bias, and finally through the sigmoid function to deal with the results.

Convolution operation

Convolution is a kind of mathematical operation that has been widely used in image processing. The result of convolution can be sorted as three modes i.e. the modes of Full, Same and Valid, which can be utilized in different occasions. For example, Valid mode is usually applied for forward propagation to facilitate the feature extraction of image, and Full mode is often employed in the back propagation to obtain the optimal weights. In the convolution operation, the operation of edge zeroing is implemented for the input image, where the layer amount of the edge can be determined according to the size of the convolution kernel [3]. The purpose of edge zeroing is to ensure the rationality of the results, i.e. the elements of the input image and the convolution kernel can be weighted and summated sequentially. Additionally, the convolution kernel should be turned around and flipped up and down as shown in Figure 3, where the kernel is actually rotated 180 degrees around the centre. It is worth noting that convolution operation can achieve sparse multiplication and parameter sharing, which can compress the dimension of the input data. In comparison with DNN, it is not necessary for CNN to provide connection weights separately for all neurons of the input data. Actually, CNN can be regarded as a common feature extraction process like most neural networks used for feature extraction.



9	8	7
6	5	4
3	2	1

0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	1	1	1	1	0	0
0	0	1	1	1	1	0	0
0	0	1	1	1	1	0	0
0	0	1	1	1	1	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

1	2	3
4	5	6
7	8	9

1	3	6	6	5	3
5	12	21	21	16	18
12	27	45	45	33	18
12	27	45	45	33	18
11	24	39	39	28	15
7	15	24	24	17	9

Fig. 5: Convolution operation

IV. RESULTS AND DISCUSSION

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

Step 1	Importing the libraries and packages
Step 2	Initializing the parameters: Batch size, Number of training epochs, Number of filters, Size of filter, Pool size
Step 3	Reading the path of input files and initialize the output folder for images after pre-processing
Step 4	Pre-processing the images for giving them as the input to the model
Step 5	Converting the images to matrix form; flattening each image into an array vector and storing them in the common image matrix

Step 6	Assigning the labels to the image classes
Step 7	Shuffling the data to prevent overfitting and generalization of training
Step 8	Separating the train data and test data
Step 9	Normalizing the images
Step 10	Defining a model and its respective layers
Step 11	Compiling the model
Step 12	Fitting the data into the compiled model, i.e., training the model using the initially defined parameters
Step 13	Plotting the Loss and Accuracy curves of the training process
Step 14	Print the Classification Report and Confusion Matrix of the training process

In this thesis work we have used datasets, FER (facial expression recognition) from still images. In the next sections we discussed both datasets in details.

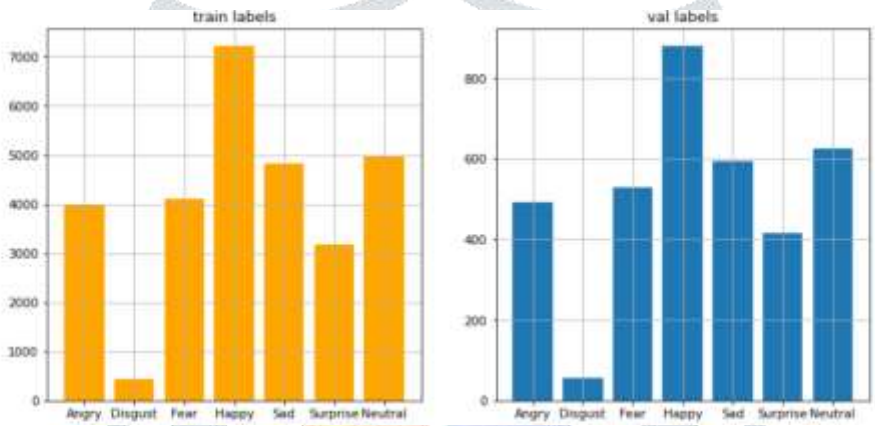


Fig. 6: Train Labels and Variable Labels Values

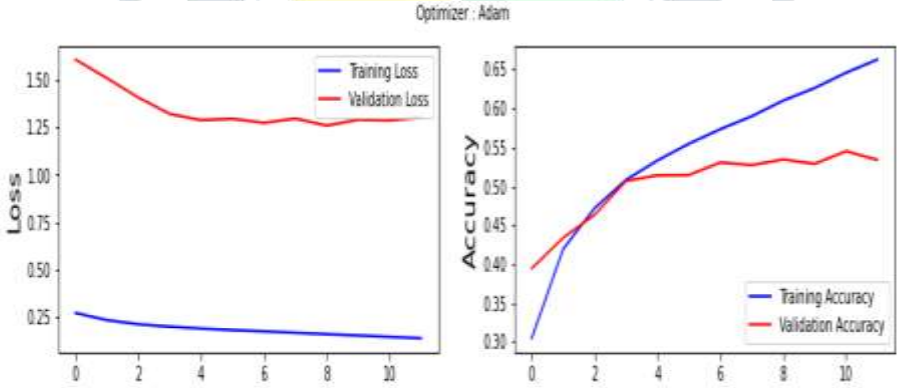


Fig. 7: Graphical Representation of Accuracy and Loss Value

V. CONCLUSION

The facial emotion recognition system presented in this research work provides a flexible facial recognition model based on mapping behavioral characteristics with physical biometric features. The physical features of the human face related to various expressions such as happiness, sadness, fear, anger, surprise and disgust are associated with geometric structures that have been restored as the basis matching template for the identity system.

REFERENCES

[1] K. M. Munim, I. Islam, M. Khatun, M. M. Karim, and M. N. Islam, ‘Towards developing a tool for UX evaluation using facial expression’, in 2017 3rd International Conference on Electrical Information and Communication Technology (EICT), 2017, pp. 1–6.

[2] J. M. Harley, ‘Measuring Emotions: A Survey of Cutting Edge Methodologies Used in Computer-Based Learning Environment Research’, in Emotions, Technology, Design, and Learning, S. Y. Tettegah and M. Gartmeier, Eds. San Diego: Academic Press, 2016, pp. 89–114.

[3] Y. Huang, F. Chen, S. Lv, and X. Wang, ‘Facial Expression Recognition: A Survey’, Symmetry, vol. 11, no. 10, Art. no. 10, Oct. 2019, doi: 10.3390/sym11101189.

[4] S. Li and W. Deng, ‘Deep Facial Expression Recognition: A Survey’, ArXiv180408348 Cs, Oct. 2018, Accessed: Nov. 16, 2019. [Online]. Available: http://arxiv.org/abs/1804.08348.

[5] O. Arriaga, M. Valdenegro-Toro, and P. Plöger, ‘Real-time convolutional neural networks for emotion and gender classification’, ArXiv Prepr. ArXiv171007557, 2017.

[6] E. Jyoti and E. A. S. Walia, ‘“A Review on Recommendation System and Web Usage Data Mining using K-Nearest Neighbor (KNN) method’, Int. Res. J. Eng. Technol. IRJET, vol. 4, no. 4, pp. 2931– 2934, 2017.

[7] Nakisa Abounasr and Hossein Pourghassem, “Facial Expression Recognition Based on Combination of Spatial-temporal and Spectral Features in Local Facial Regions”, 2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP).

[8] Mihai Gavrilescu, “Study on determining the Big-Five personality traits of an individual based on facial expressions”, The 5th IEEE International Conference on E-Health and Bioengineering - EHB 2015.

[9] Andra Adams, Marwa Mahmoud, Tadas Baltruˇsaitis and Peter Robinson, “Decoupling facial expressions and head motions in complex emotions”, 2015 International Conference on Affective Computing and Intelligent Interaction (ACII).

[10] Xunbing Shen, Wenjing Yan and Xiaolan Fu, “Recognizing Fleeting Facial Expressions with Different Viewpoints”, 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD).

[11] Juxiang Zhou Yunqiong Wanga Tianwei Xu and Wanquan Liu, “A Novel Facial Expression Recognition based on the Curvelet Features”, 2010 Fourth Pacific-Rim Symposium on Image and Video Technology.

[12] Sander Koelstra and Maja Pantic, “A Dynamic Texture-Based Approach to Recognition of Facial Actions and Their Temporal Models”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 32, NO. 11, November 2010.

[13] S.P. Aleksic and K.A Katsaggelos, “Automatic Facial Expression Recognition Using Facial Animation Parameters and Multi-Stream HMMS,” IEEE. Trans. Signal Procrssing, Supplement on Secure Media, vol. 1, pp. 3-11, March 2006.

[14] S. Koelstra, M. Pantic, and L. Patras, “A Dynamic Texture-Based Approach to Recognition of Facial Actions and Their Temporal Models,” IEEE. Trans. Pattern Analysis and Machine Intelligence, vol.32, pp. 1940-1954, 2010.

[15] A. Saha and Q.M.J Wu, “Facial Expressio Recognition Using Curvelet based Local Binary Patterns,” Proc. IEEE Int’l Conf. Acoustice Speech and Signal Procssing, pp. 2470 – 2473, March 2010.

[16] Z. Juxiang, W. Yunqiong, X. Tianwei, and L. Wanquan, “A Novel Facial Expression Recognition based on the Curvelet Features,” IEEE Conf. Image and Video Technology, pp. 82-87, 2010.

[17] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel and J. Movellan, “Recognizing facial expression: machine learning and application to spontaneous behavior,” IEEE Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 568-572 , 2005.

[18] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel and J. Movellan, “Fully Automatic Facial Action Recognition in Spontaneous Behavior,” IEEE Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 223-230, 2006.

[19] M. Yeasin and B. Bullot, “From Facial Expression to Level of Interest: A Spatio-Temporal Approach,” IEEE Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 922-927, 2004.