JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190 Vol.-14, No.-3, May-June (2019) pp 182-195 ISSN (Print) 0973-8975

Deep Learning Approach: Emotion Recognition from Human Body Movements

R. Santhoshkumar[1], M. Kalaiselvi Geetha[2]

[1,2] Department of Computer Science and Engineering, Annamalai University, Tamilnadu, India – 608002

santhoshkumar.aucse@gmail.com, geesiv@gmail.com

Corresponding Author: R. Santhoshkumar

https://doi.org/10.26782/jmcms.2019.06.00015

Abstract

Analysis of human body movements for emotion prediction is necessary for social communication. Body movements, gestures, eye movements and facial expression are some non-verbal communication method used in many applications. Among them emotion prediction from body movements is commonly used because it convey the emotional states of person from different camera view. In this paper, human emotional states predict from full body movements using feed forward deep convolution neural network architecture and Block Average Intensity Value BAIV feature. Both model can be evaluated by emotion action dataset (University of YORK) with 15 types of emotions. The experimental result showed the better recognition accuracy of the feed forward deep convolution neural network architecture.

Keywords: Emotion Recognition, Non-verbal communication, Body Movement, Human Computer Interaction (HCI), Deep Convolutional Neural Networks (DCNN), BAIV feature.

I. Introduction

In a day to day communications human beings express different types of emotions. The human communication includes verbal and non verbal communication. Sharing of wordless clues or information is called as non-verbal communication. This includes visual cues such as body language (kinesics) and physical appearance [XXIII]. Understanding human emotions is a key area of research, since recognizing emotions may provide a plethora of opportunities and applications for instance, friendlier human-computer interactions with an enhanced communication among humans, by refining the emotional intelligence [X]. Recent research on experimental psychology demonstrated that emotions are important in decision making and rational thinking. Human Emotion can be identified using body language and posture. Posture gives information which is not present in speech and facial expression. For example, the emotional state of a person from a long distance can be identified using human

posture. Hence human emotion recognition through non-verbal communication can be achieved by capturing body movement. The experimental psychology demonstrated how qualities of movement are related to specific emotions: for example, body turning towards is typical of happiness, anger, surprise; the fear brings to contract the body; joy may bring to movements of openness and acceleration of forearms brings joy; Fear and sadness brings body turning away;[III]. Emerging studies shows that people can accurately decode emotions cues from others non verbal communications and can make inference about the emotional states of others. A certain group of body actions is called as gestures. The action can be performed mostly by the head, hands and arm. These cues together and convey information of emotional states and the content in the interactions. With the support from psychological studies, identifying emotions from human body movement has plenty of applications. Suspicious action recognition to alarm security personal, human computer interaction, health care and to help autism patients is a few of the application areas of automatic emotion recognition through body cues [XIII]. In Artificial Intelligence research, a machine teaches to detect various patterns using machine learning pattern. Conventional machine-learning techniques expertise to design a feature extractor that transformed the raw data into a feature vector could detect or classify patterns in the input data. Deep learning is a dedicated form of machine learning. The technique that instructs computers to do some operation and behave like humans is done by machine learning. From the input data a machine learning task starts the feature extraction process. The features are fed to model that classifies the objects in the image. Learning feature hierarchies are produced by combining of lower level and higher level features [VII]. In deep learning model, features are automatically extracted from inputs data. Learning features automatically by several levels of abstraction. The numerous applications to machine learning techniques are ever rising at an enormous rate.

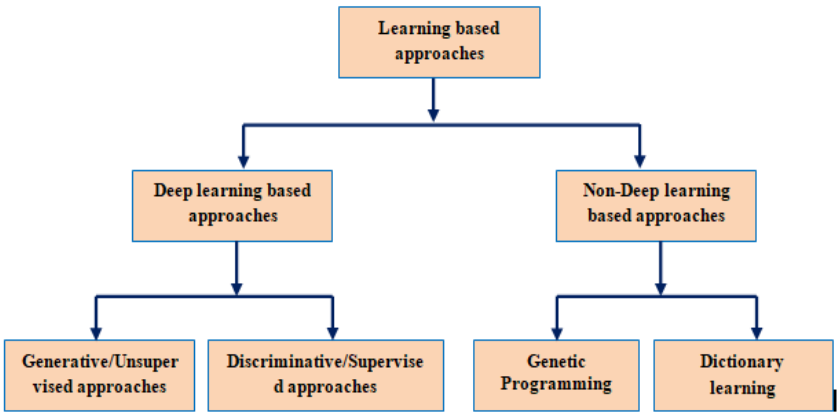

Fig 1. Learning based emotion recognition approaches

Objective of this research: This paper aims to predict emotions from human body movements with different actions using deep learning and machine learning.

Motivation of this research: In the surveillance environment, facial view is not clear when camera is too far from human. This type of issue can be rectified by capturing body movements (head, hands, legs, center of body, legs) for recognizing human emotion.

Contribution of this research: With the help of deep convolutional features the human emotion can predict easily from human body movements using feedforward deep convolution neural network (FDCNN) and recognize the emotion using BAIV feature with random forest classifier.

The following paper is as follows. Sec. II briefly summarizes the related works. Sec. III explains the proposed work. Sec. IV provides the experimental results. Finally conclusions are given and the future work is described.

II. Related Works

The computer vision applications like emotion recognition, action recognition, image and video classification can be experimented using Dictionary learning-based approaches [II]. From the large number of samples the representative vectors are learned and used in this concept. Guha and Ward [XXIV] developed a framework for human action recognition using dictionary learning methods. Based on the hierarchical descriptor the proposed method [XI] for human activity recognition outperforms the state-of-the-art methods. For a visual recognition a cross-domain dictionary learning-based method was developed [VIII]. An unsupervised model developed by Zhu and Shao for cross-view human action recognition [VI] without any label information. The coding descriptors of locality-constrained linear coding (LLC) [XXVI] are generated by a set of low-level trajectory features for each action. The Convolutional Neural Network (CNN) is the most frequently used model from the supervised category. The CNN [I] is a type of deep learning model which has shown better performance at tasks such as image classification, pattern recognition, human action recognition, hand-written digit classification and human emotion recognition. The multiple hidden layers present in the hierarchical learning model are used to transform the input data into output categories. The mapping back of different layers of CNN is called as Deconvolutional Networks (Deconvnets). The objects in the images are represented and recognized by deep CNN model. The Recurrent Neural Networks (RNNs) is the other popular model of supervised category. The skeleton-based action and emotion recognition using RNNs are developed by this author [XXV]. The five parts of human skeleton was separately fed into five subnets. The output from the subnets were combined and fed into the single layer for final demonstration. A system for automatic emotion recognition is developed using gesture dynamic's features from surveillance video and evaluated by supervised classifiers (Dynamic Time Wrapping, SVM and Naïve Bayes) [XVI]. A framework is proposed to synthesize body movements based on high level parameters and represented by the hidden units of a convolutional auto encoder [IV]. A system for recognition of affective state of person is proposed from face-and-body video using space-time interest points in video and Canonical Correlation Analysis (CCA) for fusion [XII]. The comprehensive survey of deep learning and its current applications

in sentiment analysis is discussed [XVII]. A recent works on high-performance motion data is described and relevant technologies in real time systems are proposed [XIV]. The deep learning algorithm to develop novel structure in large data sets by using the back propagation algorithm and processing images, video, speech and audio for emotion recognition are developed [XXVII]. A self-organizing neural architecture developed for recognize emotional states from full-body motion patterns [XVIII]. A system for emotion recognition on video data is developed using both CNNs and RNNs [XX].The emo FBVP database of multimodal (face, body gesture, voice and physiological signals) recordings of actors enacting various emotional expressions are predicted [XV]. A model with hierarchical feature representation for non-verbal emotion recognition and the experiments show significant accuracy improvement [XIX]. The novel design of an artificially intelligent system is proposed for emotion recognition using promising neural network architectures [V]. A novel system for Emotion Recognition in the Wild (EmotiW) is developed using hybrid CNN-RNN architecture and achieve better results over other techniques [XXII]. A new emotional body gestures is developed to differentiate culture and gender difference framework for automatic emotional body gesture recognition [IX].

III. Proposed Work

Convolutional Neural Network

Convolutional layer: Convolution has four steps

- Line up the feature and image.
- Multiply each image pixel by corresponding feature pixel.
- Add the values and find the sum.
- Divide the sum by the total number of pixel in the feature.

Which can be calculates as follows:

$$C(x_{u,v}) = \sum_{i=-\frac{n}{2}}^{\frac{n}{2}} \sum_{i=\frac{m}{2}}^{\frac{m}{2}} f_k(i,j) x_{u-i,v-j} \quad (1)$$

Where, f_k is a filter, $n \times m$ kernel size and input image is x

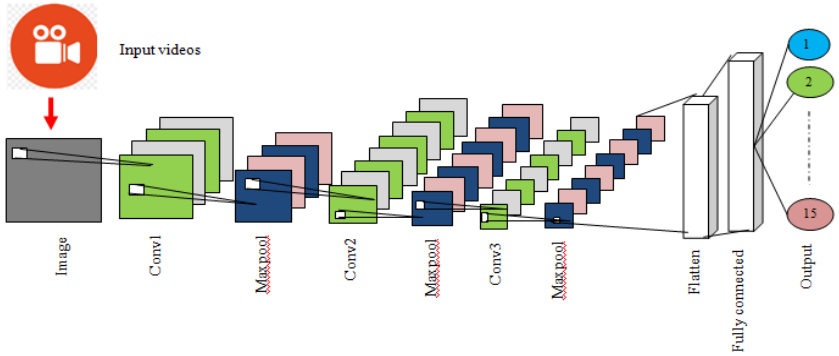

Fig 2. Convolutional Neural Network Model

Sub-sampling layers: Sub-sampling or Pooling layers shrink the map size into smaller size. The following four steps implement the pooling function:

- Pick a window size (usually 2 or 3)
- Pick a stride (usually 2)
- Move your window across your filtered images.
- The maximum value is taken from the each window.

It can be calculated by equation 2:

$$M(x_i) = Max\left\{x_{i+k,i+l}\middle||k| \leq \frac{m}{2}, |l| \leq \frac{n}{2}\, k, k\epsilon N\right\} \quad (2)$$

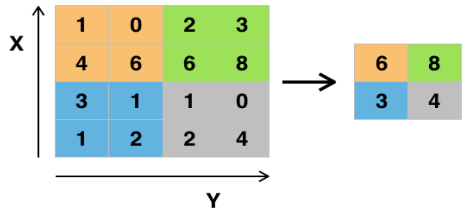

Figure 3. Example of Max-Pooling

Rectified linear unit: A rectified linear unit is an activation function while the input is below zero, the output is zero. It is calculated as follows:

$$R(x) = \max(0, x) \quad (3)$$

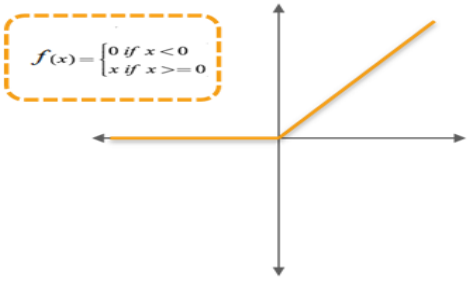

Fig 4. Rectified Linear Unit

Fully connected layer: Fully connected layer is same as neural networks in which all neurons in this layer are connected with each neuron in the previous layer. It can be calculated as:

$$F(x) = \sigma(W * x) \quad (4)$$

Softmax layer: Back propagation can be done in this layer. The networks back propagate the error and increase the performance. If *N* is a size of the input vector, so that: $S(x) : \mathrm{R} \rightarrow [0, 1]N$. It is calculated by:

$$S(x)_j = \frac{x^{xi}}{\sum_{i=0}^{N} e^{xi}} \quad (5)$$

Where $1 \leq j \leq N$

Output layer: The size of the output layer is equal to number of classes. It represents class of the input image.

$$C(x) = \{ i | \exists i \, \forall j \neq i : x_j \leq x_i \} \quad (6)$$

Feedforward Deep Convolution Neural Network (FDCNN)

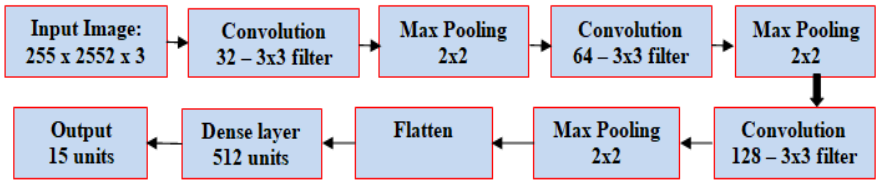

Fig 5. Architecture of FDCNN model

The input videos are converted into frames and saved in separate folder as training set and validation set. Now, the raw images are the input of first layers. The FDCNN consists of multiple convolutional layers, each of which performs the function that is discussed above. Figure 5 shows a FDCNN model. The input image is of size 150 x 150 x 3; where 3 represent colour channel. In this network the size of filter is 3 x 3 for all layers and the filter is called as weights. The multiplying of original pixel value with weight values is called sliding or convolving. These multiplications are summed and produced single number is called receptive field. Each receptive field produces a number. Finally get the feature map with size of (150 x 150 x 3). In First layer, 32 filters are applied and have 32 stacked feature maps in this stage. Then the subsampling (or max pooling) layer is reduces the spatial (feature) size of the representation with size of (75 x 75 x 32). In second layer, 64 filters are applied and have 64 stacked feature maps. Then the maxpooling layer is reduces the feature dimension to (37 x 37 x 64). In the third convolutional layer, 128 numbers of filters are applied and have 128 stacked feature maps. Then the output of the maxpooling layer is reduces the feature dimensions to (18 x 18 x 128). All max pooling layers are located with size of 2 x 2. Finally fully connected layers with 512 hidden units are placed and the output class have 15 neurons as per classes and shown the predicted emotions.

Machine Learning Approach

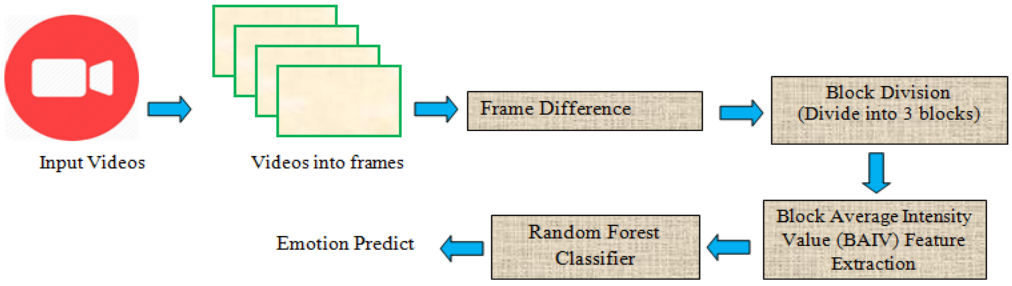

Fig 6. General Architecture of proposed model

- Frame Subtraction

Initially the input video are preprocessed and converted into frames. Frame subtraction or Frame differencing is a subtraction of frames in video sequences. It is used for change detection. The difference image is produces by two input frames time t and t + 3. The high amplitude regions shown in difference image are considered as motion regions.

$$D_k = |Int_k(i,j) - Int_{k+3}(i,j)| \quad ; \quad 1 \le i \le w, 1 \le j \le h \tag{7}$$

$D_k\ (i, j\ )$ is the difference image, $Int_k\ (i, j\ )$ is the intensity of the pixel $(i, j\ )$ in the k^{th} frame, w and h are the width and height of the image respectively. The motion region is considered as the Region of Interest (ROI). Figure 7(a), 7(b) shows the 1^{st} and 4^{th} frames of the Emotion dataset. The figure 7(c) shows the frame differencing image. Motion information $MInfo_k$ is calculated using

$$MInfo_k(i,j) = \begin{cases} 1, & if\ D_k(i,j) > t \\ 0, & otherwise \end{cases} \tag{8}$$

Where t is the threshold

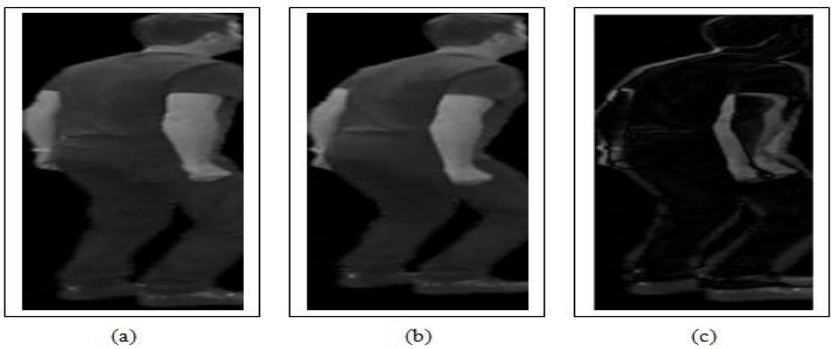

Fig7. (a) frame t, (b) frame t+3, (c) difference frame

- Feature Extraction

The input videos are preprocessed and converted into n frames. The motion information is calculated from difference image and is considered as region of interest (ROI). The difference image of size 960 x 540 is divided into three blocks B1, B2, B3 and each of size 320 x 540 pixel. Then the block having maximum intensity value is further divided into 4x5 block each of pixel size 80x108. The 20 dimensional feature vectors are extracted from the 4x5 block. The figure 8(a) and 8(b) demonstrates the block division and 20 dimensional feature extraction blocks.

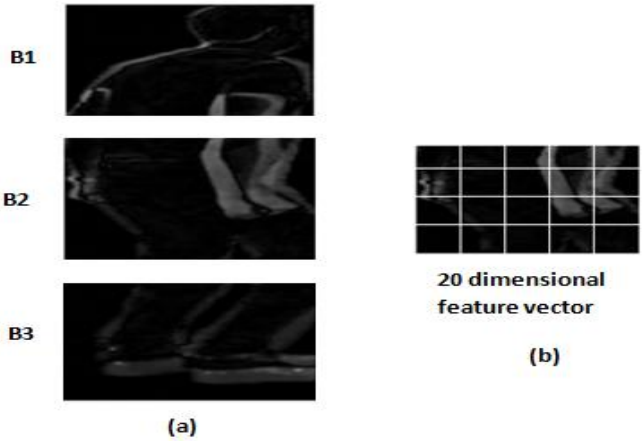

Fig 8. (a) B1, B2, B3 block division (b) 4x5 block division

IV. Performance Evaluation Metrics

Dataset

Emotion dataset (University of York) containing five different emotions (happy, angry, sad, untrustworthy and fear) performed by 29 actors. The sequences were taken over static (black) background with the frame size of 1920 × 1080 pixels at a rate of 25 fps.

Fig 9. Example frame of 5emotions and 3 actions

All the experiment were implemented by Windows 10 operating system with Intel core i5 3.3 Ghz, processor with Anaconda Python and Jupyter notebook. The dataset is trained with batch size 64 and 10 epochs.

Random forest classifier

The Random Forest (RF) algorithm is one of the supervised classification algorithms [XXI]. The name itself described which is to create a forest by someway and make it random. There is a direct relationship between the number of trees in the forest and the results. When increase the number of trees, the accuracy of the result can increase. The two stages of RF algorithms are creation and prediction. The creation of Random Forest algorithm is as follows:

- Select “K” features from total “m” features randomly where k << m
- From the “K” features, calculate the node “d” using the best split point
- Using the best method, split the node into daughter nodes
- Repeat the a to c steps until “l” number of nodes has been reached
- Build forest by repeating steps a to d for “n” number times to create “n” number of trees

The prediction of Random Forest algorithm is as follows:

- Takes the test features and use the rules of each randomly created decision tree to predict the outcome and stores the predicted outcome (target)
- Calculate the votes for each predicted target
- Consider the high voted predicted target as the final prediction from the random forest algorithm.

Evaluation Metrics

The confusion matrix is a table with actual classifications is columns and predicted ones are Rows.

- True Positives (TP) - when the data point of actual class and predicted was (True).
- True Negatives (TN) - when the data point of actual class and predicted was (False).
- False Positives (FP) - when the data point of actual class (False) and the predicted is (True).
- False Negatives (FN) - when the data point of actual class (True) and the predicted is (False).

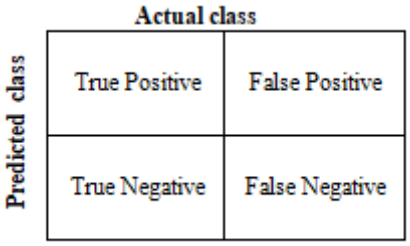

Fig 10. Confusion matrix

The performances evaluation of proposed work can be calculated using Accuracy, Recall, F-Score, Specificity and Precision. Accuracy in classification problems is the number of correct prediction made by the model over all kinds prediction made, which can be calculated using equation 9.

$$Accuracy = \frac{tp+tn}{tn+fp+tp+fn} \tag{9}$$

Recall provides how extraordinary an emotion is recognized accurately.

$$Recall \ = \frac{tp}{tp+fn} \tag{10}$$

The symphonies mean of Precision and Recall is called as F-score.

$$F-Score = 2\frac{Precision \ \ X \ Recall}{Precision \ +Recall} \tag{11}$$

Specificity shows an evaluation of how great a strategy is recognizing negative emotion accurately.

$$Specificity = \frac{tn}{tn+fp} \tag{12}$$

At last, Precision shows the proportion of classification, which can be calculated by equation (13).

$$Precision = \frac{tp}{tp+fp} \tag{13}$$

Where, tp and tn are the quantity of true positive and true negative prediction of the class and fp and fn are the quantity of false positive and false negative expectations.

V. Results and Discussions

The table 1described the correlation among happy jump and the untrustworthy jump and the angry jump which point out those images are somewhat closer to each other. The performance of this network good in predict happy walk, sad walk, sad sit, happy sit than other emotions. Which point out learning those category of emotions is perfect than other emotions.

Table 1: Individual accuracy (%) for fifteen class emotions

	AJ	AS	AW	FJ	FS	FW	HJ	HS	HW	SJ	SS	SW	UJ	US	UW
AJ	95	0	0	1	0	0	2	0	0	0	0	0	2	0	0
AS	0	96	0	0	1	0	0	2	0	0	1	0	0	0	0
AW	0	0	96	0	0	1	0	0	1	0	0	0	0	0	2
FJ	2	0	0	94	0	0	1	0	0	2	0	0	1	0	0
FS	0	1	0	0	95	0	0	1	0	0	2	0	0	1	0
FW	0	0	3	0	0	95	0	0	2	0	0	0	0	0	0
HJ	0	0	0	1	0	0	96	0	0	2	0	0	1	0	0
HS	0	1	0	0	1	0	0	97	0	0	0	0	0	1	0
HW	0	0	0	0	0	1	0	0	94	0	0	3	0	0	2
SJ	4	0	0	1	0	0	1	0	0	94	0	0	0	0	0
SS	0	3	0	0	0	0	0	1	0	0	95	0	0	1	0
SW	0	0	4	0	0	2	0	0	0	0	0	92	0	0	2
UJ	0	0	0	2	0	0	3	0	0	4	0	0	91	0	0
US	0	3	0	0	2	0	0	1	0	0	2	0	0	92	0
UW	0	0	2	0	0	5	0	0	1	0	0	1	0	0	91

Note: AJ=AngryJump, AS=AngrySit, AW=AngryWalk, FJ=FearJump, FS=FearSit, FW=FearWalk, HJ=HappyJump, HS=HappySit, HW=HappyWalk, SJ=SadJump, SS=SadSit, SW=SadWalk, UJ=UntrustworthyJump, Us= UntrustworthySit, UW=UntrustworthyWalk

Table 2. Performance Measure of Emotion dataset

	Precession		Recall		F-Measure	
Model	**DL**	**ML**	**DL**	**ML**	**DL**	**ML**
AJ	0.937	0.891	0.928	0.821	0.943	0.881
AS	0.881	0.854	0.897	0.885	0.939	0.877
AW	0.912	0.825	0.921	0.913	0.921	0.901
FJ	0.901	0.862	0.897	0.886	0.909	0.891
FS	0.891	0.869	0.932	0.924	0.919	0.876
FW	0.895	0.883	0.891	0.883	0.927	0.879
HJ	0.884	0.873	0.935	0.927	0.915	0.913
HS	0.862	0.854	0.899	0.879	0.929	0.902
HW	0.883	0.839	0.928	0.911	0.931	0.911
SJ	0.931	0.930	0.918	0.849	0.947	0.886
SS	0.852	0.844	0.887	0.879	0.955	0.897
SW	0.892	0.882	0.931	0.922	0.921	0.879
UJ	0.933	0.924	0.918	0.821	0.913	0.880
US	0.891	0.887	0.890	0.881	0.939	0.878
UW	0.892	0.831	0.910	0.928	0.915	0.909

Note: DL-Deep learning, ML-Machine learning

VI. Conclusion

In this paper, proposed FDCNN and 20 dimensional BAIV feature for predicting human emotions from body movements on videos. One model is representing deep convolutional features to extract saliency information at multiple scales and another model extract BAIV feature and fed to the Random forest classifier for emotion prediction. The proposed methods are evaluated on challenging benchmarks Emotion dataset (University of York). The performance of FDCNN model is better than BAIV features. In Future work, the author aims to develop a novel system to recognize the emotions of children with autism spectrum disorder (ASD) using body movements (head, L-hand, R-hand, center of body).

References

I. A.Krizhevsky, I. Sutskever, and G. E. Hinton, (2014),"Imagenet Classification With Deep Convolutional Neural Networks," *In Advances in neural information processing systems*, pp. 1097–1105.

II. D.Tran, L. Bourdev, R. Fergus, L.Torresani and M. Paluri, (2015), "Learning Spatiotemporal features with 3d Convolutional networks", *IEEE International Conference on Computer Vision (ICCV)*, pp. 4489-4497.

III. Damel Rucha, Gurjar Aditya, Joshi Anuja, Nagre Kartik, (2015), "Human Body Skeleton detection and Tracking", *International Journal of Technical Research and Applications*, Volume 3, Issue 6, pp.222-225.

IV. Daniel Holden, Jun Saito, Taku Komura. (2016) "A Deep Learning Framework for Character Motion Synthesis and Editing" *SIGGRAPH '16 Technical Paper, July 24 - 28, Anaheim*, CA, ISBN: 978-1-4503-4279-7/16/07.

V. Enrique Correa, Arnoud Jonker, Michael Ozo, Rob Stolk. (2016) "Emotion Recognition using Deep Convolutional Neural Networks"

VI. F. Zhu and L. Shao, (2014),"Weakly-Supervised Cross-Domain Dictionary Learning for Visual Recognition," *International Journal of Computer Vision*, Vol. 109, No. 1-2, pp. 42–59.

VII. F. Zhu and L. Shao, (2015),"Correspondence-Free Dictionary Learning for Cross-View Action Recognition," *In ICPR*, pp. 4525–4530.

VIII. F. Zhu, L. Shao, J. Xie, and Y. Fang, (2016),"From Handcrafted to Learned Representations for Human Action Recognition: A Survey," *Image and Vision Computing*.

IX. Fatemeh Noroozi, Ciprian Adrian Corneanu, Dorota Kami´nska, Tomasz Sapi´ nski, Sergio Escalera, and Gholamreza Anbarjafari (2015) "Survey on Emotional Body Gesture Recognition" Journal of IEEE Transactions on Affective Computing.

X. Gavrilescu, M., (2015) "Recognizing emotions from videos by studying facial expressions, body postures and hand gestures", *23rd Telecommunication fourm TELFOR*, pp. 720-723.

XI. H.Wang, C. Yuan,W. Hu, and C. Sun,(2012), "Supervised Class-Specific Dictionary Learning for Sparse Modeling in Action Recognition," *Pattern Recognition*, Vol. 45, No. 11,pp. 3902–3911.

XII. Hatice Gunes, Caifeng Shan, Shizhi Chen, YingLi Tian. (2015) "Bodily Expression for Automatic Affect Recognition. Emotion Recognition: A Pattern Analysis Approach" Published by John Wiley & Sons, Inc.

XIII. Hazel Rose Markus, Shinobu Kitayama.(1991) "Culture and the self: Implementations for cognition, emotion, and motivation" *Psychological Review*, pp. 224-253.

XIV. Heike Brock. (2018) "Deep learning - Accelerating Next Generation Performance Analysis Systems" *12th Conference of the International Sports Engineering Association, Brisbane, Queensland, Australia*, pp. 26–29.

XV. Hiranmayi Ranganathan, Shayok Chakraborty, Sethuraman Panchanathan.(2017) "Multimodal Emotion Recognition using Deep Learning Architectures" http://emofbvp.org/

XVI. J. Arunnehru, M. Kalaiselvi Geetha. (2017) "Automatic Human Emotion Recognition in Surveillance Video" *Intelligent Techniques in Signal Processing for Multimedia Security, Springer-Verlag*, pp. 321-342.

XVII. Lei Zhang, Shuai Wang, Bing Liu. (2018) "Deep Learning for Sentiment Analysis: A Survey" https://arxiv.org/pdf/1801.07883.

XVIII. Nourhan E, Pablo B, Parisi, Stefan Wermter, (2017),"Emotion recognition from body expressions with Neural Network Architecture", *Algorithm and Learning, HAI 2017*, pp. 143-149.

xix. Pablo Barros, Doreen Jirak, Cornelius Weber, Stefan Wermter. (2015) "Multimodal emotional state recognition using sequence-dependent deep hierarchical features" *Neural Networks*. 72, pp. 140–151.

xx. Pooya Khorrami, Tom Le Paine, Kevin Brady, Charlie Dagli, Thomas S. Huang. (2017) "How Deep Neural Networks Can Improve Emotion Recognition on Video Data" https://arxiv.org/pdf/1602.07377.pdf.

xxi. Prinzie, A., Van den Poel, D., (2012), Random Forests for multiclass classification: Random MultiNomial Logit. *Expert Systems with Applications*. Vol.34, 3, pp.1721–1732.

xxii. Samira Ebrahimi Kahou, Vincent Michalski, Kishore Konda, Roland Memisevic, Christopher Pal. (2015) "Recurrent Neural Networks for Emotion Recognition in Video" *ICMI 2015, Seattle, WA, USA.*

xxiii. Shirbhate Neha, Talele Kiran, (2016), "Human Body Language Understanding for Action detection using Geometric Features", *2nd International Conference on Contemporary Computing and Informatics, IEEE*, pp.603-607.

xxiv. T. Guha and R. K.Ward, (2012),"Learning Sparse Representations for Human Action Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 8, pp. 1576–1588.

xxv. Y. Du,W.Wang, and L.Wang, (2015),"Hierarchical Recurrent Neural Network for Skeleton Based Action Recognition," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1110–1118.

xxvi. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel,(1989), "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural computation*, Vol. 1, No. 4, pp. 541–551.

xxvii. Yann LeCun, Yoshua Bengio, Geoffrey Hinton.(2015) "Deep learning" *Nature*, Vol. 521, pp. 436-444.