Estimating Attention of Faces due to its Growing Level of Emotions

Ravi Kant Kumar*, Jogendra Garain, Dakshina Ranjan Kisku and Goutam Sanyal
Department of Computer Science and Engineering
National Institute of Technology
Durgapur, India
E-mail: {vit.ravikant, jogs.cse, drkisku, nitgsanyal}@gmail.com

Abstract

In the task of attending faces in the disciplined assembly (Like in examination hall or Silent public places), our gaze automatically goes towards those persons who exhibits their expression other than the normal expression. It happens due to finding of dissimilar expression among the gathering of normal. In order to modeling this concept in the intelligent vision of computer system, hardly some effective researches have been succeeded. Therefore, in this proposal we have tried to come out with a solution for handling such challenging task of computer vision. Actually, this problem is related to cognitive aspect of visual attention. In the literature of visual saliency authors have dealt with expressionless objects but it has not been addressed with object like face which exploits expressions. Visual saliency is a term which differentiates "appealing" visual substance from others, based on their feature differences. In this paper, in the set of multiple faces, 'Salient face' has been explored based on 'emotion deviation' from the normal. In the first phase of the experiment, face detection task has been accomplished using Viola Jones face detector. The concept of deep convolution neural network (CNN) has been applied for training and classification of different facial expression of emotions. Moreover, saliency score of every face of the input image have been computed by measuring their 'emotion score' which depends upon the deviation from the 'normal expression' scores. This proposed approach exhibits fairly good result which may give a new dimension to the researchers towards the modeling of an intelligent vision system which can be useful in the task of visual security and surveillance.

1. Introduction

Emotions with faltering in the facial muscles are called facial expressions [1]. It reveals the current attitude of a person and it facilitates us to make non-verbal communication among the people. Moreover, different facial expression also helps to judge the different mindset and feeling [2] of a person at that moment. Facial expression plays an imperative role in non-verbal communication as well as to predicting the behavior of the person. During a group discussion, our attention automatically goes towards those participants who put more stressed on his words or talk in a sentimental or emphatic voice. Same phenomenon occurs with the non-verbal visual communication. The face reflecting the higher expression of a particular emotion draws more attention [3, 4] in the discussion. A particular object (It also may be face), which gives us more visualization is consider as a salient object and this phenomenon is called visual saliency [4, 5]. Computer modeling with the salient facial expressions might be used in numerous applications like; Human Behavior Analyzer [6], Surveillance System [7], Mental Disease Diagnosis [8] etc.

There are several research papers available on visual saliency and attention. But in these literatures, saliency modeling has been developed for the objects only. Some of them are:

Graph-Based Visual Saliency [9], Degree Centrality Based Model [5]. Frequency-tuned Salient Region Detection [10], Minimum Barrier Salient Object Detection [11], Context-Aware Saliency [12] and Global Contrast Based Salient Region Detection [13]. But none of these approach deals with the high priority object like faces, where expression does matters for guiding our focus.

However, a few amounts of work have been accomplished where saliency has been modeled with inclusion of faces. These methods are: Saliency in crowd [14], Anomaly Detection in Crowded Scenes [15], Attention capture by faces [16], Saliency map augmentation with facial detection [17], Enlighten the effect of neighbor faces in the crowd [18], Saliency based intelligent camera for enhancing viewer's attention [19]. But these models only include the basic facial features and not incorporating the attention effect because of facial emotions and expressions. Therefore, we have proposed a novel technique to determine facial saliency which occurs because of face expression and emotions. In this work, relative saliency of faces has been determined based on its emotional score and their respective locality in the scene.

The task of emotion detection has been accomplished through training and classification using deep convolution neural network. For handling such complex task of emotion classification, huge features are required to train the model. Convolution Neural Network is capable to extract the enormous features, further it can generate another set of features after various combinations of features and classify the emotions using advanced machine learning methods like Dropout.

Further, the paper is organized as: In section 2, Database description has been mentioned. Complete system architecture has been briefly explained in section 3. In section 4, emotion score has been computed using deep CNN. Section 5, described about the preparation of ground truth. Mathematical modeling has been proposed in section 6. In section 7, saliency map has been generated based on the estimated emotion score.

Experimental has been validated in section 8. Finally, section 9 draws the concluding remarks.

2. Database Description

In order to train and test our model for the various emotions, we have used two databases. System has been train using FER-2013[20] database. It contains 28,709 grayscale images, each having size of 48x48. Before using these samples for the training, we have explored the valid sample as a preprocessing task, which is described in section 3.1. FER-2013 dataset consists six types of emotions (Angry, Disgust, Fear, Happy, Sad, and Surprise). It exhibits emotions in low resolution images taken from arbitrary distances under unconstraint condition. Therefore, we have used this database for train the system. As the system has been trained with the challenging database, it can easily classify the emotion in standard conditions.

Figure 1: Sample FER-2013[20] Database.

Testing has been performed over Cohn-Kanade Facial Expression Database (CK+) [21]. It contains 486 face images of 97 subjects. For every subject, there are various levels of emotions sequence which starts from a neutral expression and ends with its peak expression. The increasing levels of emotions are very clearly visible because images are obtained through high resolution camera captured from the same distance. Since image quality of CK+ database [21] is better than FER-2013[20], therefore, testing performance of classification of emotions is good enough.

Figure 2: Sample CK+ [21] Database

3. System Architecture

The entire architecture of our proposed system has been described in Fig 3. It mainly consists of four phases namely Pre-processing phase (Face Detection), Training phase, testing phase and finally, Saliency phase.

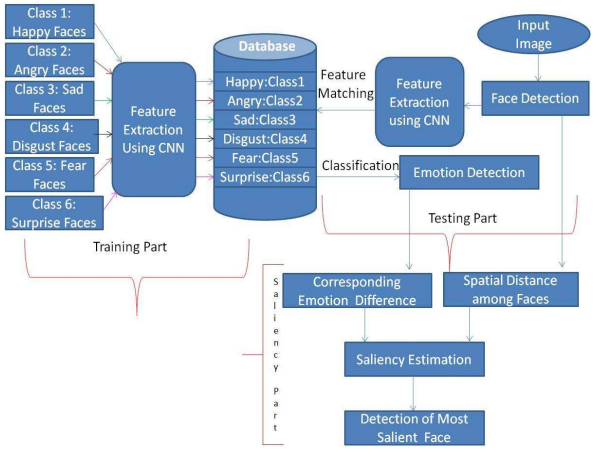

Figure 3: Proposed System Architecture

3.1. Face Detection as Preprocessing

We humans first detect a face then only recognize its emotion. It means, face detection is the pre-process of emotion detection. Therefore, the same procedure has been followed in this proposed approach. In order to getting the salient face among the faces having the same emotions with its different level of expression, there is a requirement of face detection task. One of the most popular Viola Jones Face Detector [22] has been applied for detecting the faces. It extracts Haar-Like Features [23] from a face. Haar-Like Features are the various masks which extract various features for the object categorization. Extraneous redundant features are cast-off by AdaBoost algorithm [22, 23]. If a sub-window carried out similar features as pre-define facial features, set by the Cascaded Classifier [22], at any stage is regarded as a face.

Figure 4: Detected Faces using Viola Jones Algorithm [22]

In the sub-window area, a bounding box is created around the face in the image. The area covered inside the bounding box is cropped and resized into 48x48 pixels. After pre-processing, the dataset consists of 11,246 images of the 7 emotions of which 1456 are angry, 240 are disgust, 1414 fear, 3235 happy, 1304 sad, 1362 surprise and 2235 are neutral. All the images are of frontal face.

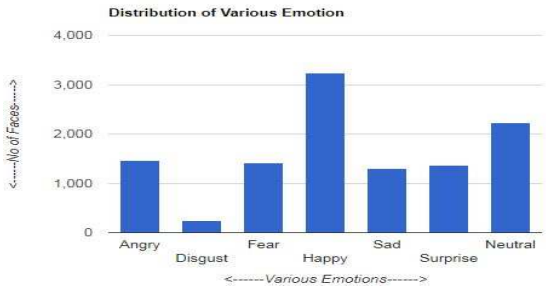

Figure 5: Distribution of Emotions after Pre-processing

3.2. Training Phase

In the training part, feature extraction, its various combinations and training classification have been done using deep convolution neural network [24]. In this process, supervised learning approaches have been applied over huge number of face images. FER-2013 database contains six standard categories of emotions of 7700 face images. In the Convolutional layer, pixels values of the input layers are getting convolve with some certain appropriate weights. A number of random filters are convolving with image to extract features for training classification. 'Training classification' is nothing but associating an emotion label to the all the obtained feature vectors of the corresponding class of emotion. The detail working of convolutional neural network (CNN) has been described in section 4.

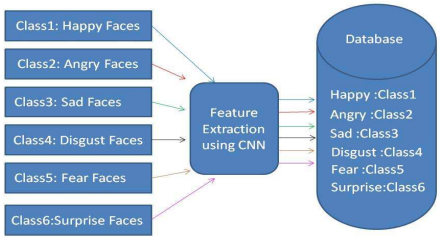

Figure 6: Training Architecture

3.3. Testing Phase

In our proposed work, there was a requirement of a database which exhibits various level of same emotion for the same subject. As, Cohn-Kanade Facial Expression Database (CK+) [21] meets with all these requirement, therefore, testing has been performed with it. As like training phase, in the testing phase, feature extraction has also been done using the same convolutional neural network. Classification of emotion is done after matching of these features with the feature vectors of the trained database. Extracted feature values (Available just before the output layer) are compared with feature values available in the training feature set. If the feature value of test image matched with the training feature score, within specified threshold range, the corresponding emotion label are allotted to the test image. Moreover, emotion scores are also computed based on the probability of belongingness to all the 6-emotion training class.

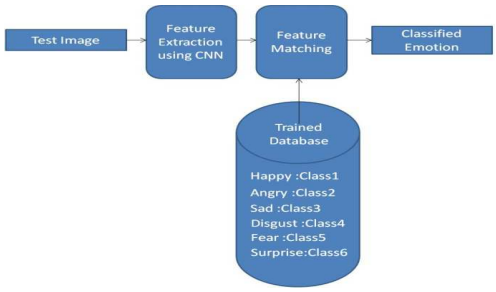

Figure 7: Testing Architecture

3.4. Saliency Phase (Proposed)

In this phase, saliency of a face is determined based on strength of a particular emotion appear over it. Relative saliencies of faces are calculated on the basis of the difference of emotion score and their proximity difference. The task of emotion detection has been accomplished through training and classification using deep convolution neural network. In all 6 emotion scores of a face, the maximum scores signify the effective emotion. Therefore, maximum score has been considered for producing the saliency map. In this work, proposed saliency map has been generated by inclusion of emotion scores. The Frequency-tuned salient region detection method [10] is a well-known novel state of art method for finding the salient location using low-level features and verified with the ground truth. This technique [10] presents outstanding results for determining object saliency but unable to figure out the attention for facial expressions and emotions. Therefore, we have implemented the emotion effect in the method [10]. The saliency architecture has been shown in Figure 8.

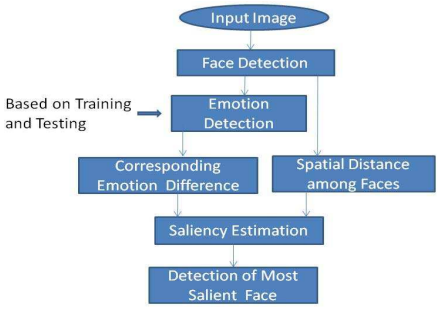

Figure 8: Saliency Architecture

4. Emotion Estimation Using Deep Convolution Neural Network

In these days, deep learning is a hot topic for solving various training and classification problem. A convolution neural network (CNN) [24, 25] is the backbone of designing the deep learning architecture. We humans are visualizing and deeply learning about the various expressions and emotions from our childhood. Hence, we are enough capable to recognize various emotion very easily. In the same way, computer vision system also requires huge training over large dataset to recognize the emotions in human faces. Therefore, in the proposed works emotions recognition has been completed using deep convolution neural network (CNN). The performance of neural network depends on several parameters like random weights, activation function, input data for training and testing, hidden layers and network architecture. The convolution neural network receives input images through the input layer. In the hidden layers, it extracts features and forward to the subsequent layers for extracting various combinations of features and further uses in the training and testing classification process.

In order to recognizing the six standard emotions, nine layered CNN architectures have been employed, which is shown in the Figure 9.

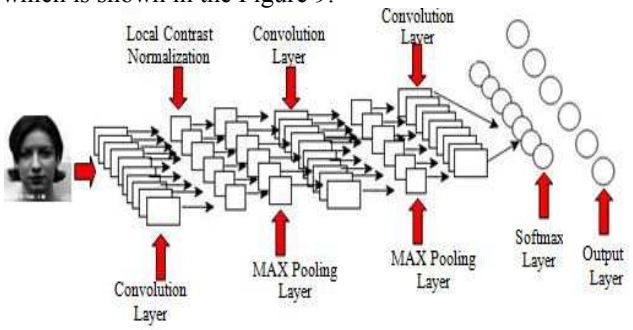

Figure 9: Deep Convolutional Neural Network Architecture

The working process of configured architecture of Convolutional Neural Network [24, 25] has been explained below:

'Keras' [26] is a deep learning python library which facilities 'Tensorflow' [26] in the backend. It provides an efficient and fast implementation of neural network. This network (Fig 9) has been framed using ConvNet architectures [26] which takes inputs as images, extract various features and combined these features automatically in the subsequent hidden layers. Finally, in the output layer, it classifies the various emotions.

Input layer receives the pixels values from the input images (with dimension w x h x c) where w, h and c are the corresponding width, height and no of colour channels. All input images are resized into 48x48 for better compatibility. Gray scale images have been used in our experiment, therefore, dimension of images is considered to be 48x48x1, where 1 signify the gray scale images.

Every pixel location is considered to be a neuron. Convolutional layer computes the dot product of weights and the feature value of the input image location (neurons). The various filters assigned with some random weight are convolved with the image. It generates various feature maps corresponds to the pixels location. In these feature maps edges and the orientations appears which signify how the pixels locations are improved. This result in (w x h x f), where f is the number of filters used.

Next, Pooling layer do the down sampling to reduce the redundancy in the feature maps generated by convolutional layers. It also reduces the computational time for the further process as it diminishes the dimensions of the map by a factor of window size. Hence, in the resultant feature maps, only the pixels having the maximum values are retained.

Softmax Layer transforms the features through those layers which are connected with trainable weights. This layer combines the refined features in the entire image. Sometimes it causes the problems of over fitting. But it can be overcome by adding a dropout layer during the training process.

Dense or Fully Connected layer is entirely linked with the output of the preceding layer. It is used in the pre-final stage of CNN for connecting with the output layer to forming the desired number of outputs. It transforms the features through layers connected with trainable weights. This layer recognizes the sophisticated features in the output image. Sometimes it grows into prone to over fitting. This is condensed by accumulating a dropout layer which randomly chooses a portion (Normally less than 50%) of image to set their weights to zero during training. Output layer is connected to the previous fully connected layer and outputs the required classes or their probabilities. Since, in human some emotions are generally a consolidation of emotions which is computed by probability of each emotion. This is achieved by using Softmax layer in the network.

5. Preparing Ground Truth

In this work, the saliency of the faces has been determined based on their emotions. In general day to day life, during conversations we observe that our attention goes towards the faces which are expressing higher degree of emotion rather than a normal or lesser one. By keeping the same concept in mind, ground truth data has been prepared considering the highest level of emotion as the most salient.

Figure 10: Ground Truth based on Highest Level of Emotion

6. Proposed Mathematical Formulation for Estimating Saliency Score due to Emotion

In general, relative saliency between two faces is determined by their contrast difference. 'Contrast difference' is nothing but their feature difference. In degree centrality model [5] and graph-based saliency model [9], relative distances among the objects have also been taken into account. Grasping this concept in mind, we have proposed a new mathematical model for computing saliency of a face due to its own emotion with respect to the emotions of its surrounding faces as:

The saliency weight S_{ij} between any pair of faces 'i' and 'j' is determine by 'emotion difference' modulated with 'positional proximity' between them.

$$S_{ij} = (E_i - E_j)\, e^{-D_{S_{ij}}^2 / 2\sigma^2} \quad (1)$$

Where E_i and E_j are the emotion score of the two faces i and j. This is computed by deep convolution neural network (described in section 3 and 4). The spatial distance $D_{S_{ij}}$ between any faces i and j is calculated as the Cartesian distance between the centers of the corresponding faces.

In general, saliency of face 'i', (i.e. S_i) due to its surrounding 'n' numbers of faces, is computed as:

$$S_i = \sum_{j=1}^{n} (E_i - E_j)\, e^{-D_{S_{ij}}^2 / 2\sigma^2} \quad (2)$$

7. Proposed Mathematical Formulation for Estimating Net Saliency and Map Generation

The proposed saliency map has been developed using frequency tuned technique [10] amalgamated with emotion saliency score of corresponding faces. Saliency emotion score have been calculated by Emotion score of using CNN. The proposed saliency map has been generated by fusing the emotion saliency score of the corresponding faces in the method [10], which has been mentioned in equation (3).

The net saliency weight NS_{ij} between any pair of faces 'i' and 'j'

$$\left.\begin{aligned} &NS_{ij} = S_{FrequencyTuned}(Face_i) + S_{ij} \\ &\text{i.e. } NS_{ij} = S_{FrequencyTuned}(Face_i) + \\ &(E_i - E_j)\, e^{-D_{S_{ij}}^2 / 2\sigma^2} \end{aligned}\right\} \quad (3)$$

In general, Net saliency of face 'i', (i.e. NS_i) due to all its surrounding 'n' numbers of faces is computed as:

$$\left.\begin{aligned} &NS_i = S_{FrequencyTuned}(Face_i) + S_i \\ &\text{i.e. } NS_i = S_{FrequencyTuned}(Face_i) + \\ &\sum_{j=1}^{n} (E_i - E_j)\, e^{-D_{S_{ij}}^2 / 2\sigma^2} \end{aligned}\right\} \quad (4)$$

The entire steps for computing saliency of faces have been described below:

7.1. Face Detection

In order to find the emotions and saliency of faces in the input image, there is a requirement of face detection task. The faces are detected using Viola Jones algorithm [22].


Figure 11: Detected Face using Viola Jones algorithm [22]

7.2. Finding Center

For computing the spatial proximities among the faces, center of faces has been calculated using normal geometry.


Figure 12: Ground Truth for various Intense Emotion

Center coordinates of faces are depicted in Table 1.

Table 1. Center Coordinates of Faces

Face No	1	2	3
X-Co	201.49	531.95	859.72
Y-Co	119.66	125.040	119.66

7.3. Emotion Score

Now, emotion of all the detected faces have been determined by trained convolutional neural network. The detailed about training, testing and emotion computation have been briefly described in the sections (3 and 4). Emotion scores computed using trained CNN (Convolution Neural Network) have been mentioned in Table 2.

Table 2. Emotion Score of Faces

Face No	1	2	3
Emotion Score	0.03	57.1	99.91

7.4. Saliency Estimation

Based on emotion score (Table 2), spatial distances among faces (Computed using Table 1), saliency values have been estimated using proposed mathematical formulation described in Equation (2).

Table 3. Saliency Score of Faces

Face No	1	2	3
Saliency	0.0002	**1.0000**	0.4505

Based on obtained saliency scores, the most salient face has been depicted in Figure 13.

Figure 13: Most Salient Face Bounded in Green Box

7.5. Saliency Map Generation

The proposed saliency map has been generated by incorporating the emotion scores of the faces using equations 1, 2, 3 and 4. A face may be salient due to presence of high level features (Facial Expressions, Emotions) as well as the low-level features (Color, Intensity, Texture etc.). The concept of well-known frequency-tuned technique [10] has been taken into account for dealing with the low-level features. High level feature map has been generated by considering the emotion score of the faces. The obtained saliency score using our proposed technique is depicted in Table 3.

The reason behind choosing the method [10] for extracting low level features, because in this technique, authors have filled the drawback of many state of art methods [5, 9, 27, 28] for computing saliency. These methods depict salient regions with low resolution, poorly borders and also, they are expensive for extracting salient features. Moreover, some other methods like [12] highlight the object edges instead of highlighting the whole object, but they consume entire spatial frequency of the input image.

In the frequency-tuned technique [10], authors calculated the spatial frequencies and explored a frequency-tuned technique for computing saliency based on the center-surround difference with intensity features.

In frequency-tuned technique [10], saliency map S for an image I having dimensioned (width = W and height = H pixels) is formulated as per Equation 5.

$$S(x,y) = | I_{\mu} - I_{w_{hc}}(x,y) | \quad (5)$$

Where I_{μ} is the mean pixel value and $I_{w_{hc}}$ indicates the Gaussian blurred version of the original image *I*. Gaussian blurred version is taken for removing the deep texture details and noises.

In order to inclusion of high level features to the proposed saliency map, emotion scores have been computed by proposed deep CNN, mentioned in section 4. The proposed saliency map has been generated by fusing the low-level features map using method [10], with the corresponding saliency score obtained from the CNN. The proposed saliency formulation has been mentioned in equation (3).

Figure 14: Top: Saliency Map using [10], Middle: Saliency Map due to emotion, Bottom: Proposed Saliency Map after Fusing Emotion Score with method [10]

8. Experimental Validation and More Results

Experiment has been conducted on various input image set, reflecting different level of expression of a particular emotion of the same face. Several input images have been created from CK+ dataset [21], after selecting various level of same emotion of the same person. Salient face for all the series of emotion has been computed using our proposed technique. Some of the experimented results have been shown in Figure 15, where salient face of each of the series (i.e. happy, disgust, surprise, angry, fearful and sad) is highlighted in green box.

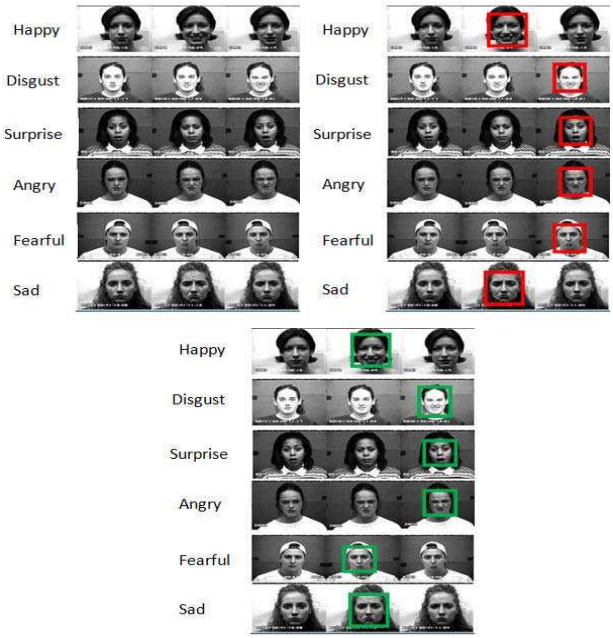

Figure 15: Up Left: Input Image in various Emotions in different level, Up Right: Ground Truth, Down: Most Salient Face indicated in Green Box

For the given set of above input images having different level of same emotion, the respective emotion score of the faces obtained using CNN have been described in the following Table 4.

Table 4. Normalized Emotion Score of Faces

Face No→	1	2	3
Happy	0.0002	**1.0000**	0.4505
Disgust	0.0002	0.4890	**1.0000**
Surprise	0.9095	0.0005	**1.0000**
Angry	0.6426	0.6983	**1.0000**
Fearful	0.5583	**1.0000**	0.4873
Sad	0.3365	**1.0000**	0.8831

Finally, saliency scores have been computed based on obtained emotion score (Table 3), spatial distances among faces (Table 1), using equation (2).

Table 5. Saliency Score of Faces

Face No→	1	2	3
Happy	0.03	**99.91**	57.1
Disgust	0.01	22.51	**58.57**
Surprise	0.05	84.1	**97.32**
Angry	79.3	85.91	**96.98**
Fearful	35.11	**50.54**	31.96
Sad	36.82	85.75	**95.8**

This research aim is to finding the most salient face due to presence of different level of same emotion in a same face. The ground truth has been prepared by considering the highest level of emotional face as the most salient as per our perceptions (Fig 10). The most salient face obtained through our approach is matching with the ground truth for most of the cases. In the ground truth, for the 'fearful input image' Face No 3 is the most salient face. But using our proposed work, Face No 2 is the most salient. It happens because, for Face No 2, emotion scores (Table 4) has found to be highest using trained deep convolution neural network. As in our approach, saliency score has been obtained using these emotional scores therefore, obtained saliency score (Table 5) has also been affected.

For the sad input image (Last series), 'Face No 2' is the most salient as per the ground truth (Fig 10). Highest emotion score (Table 5) has been found for 'Face No 2' through trained deep convolution neural network. These emotion score have been used in computing proposed saliency score. But, 'Face No 3' has achieved the highest saliency value. Here it happens because; in the proposed work saliency has been computed based on relative emotion score as well as their relative spatial locations. Therefore, obtained resultant values determined 'Face No 2' as the most salient.

The most salient faces obtained using our proposed approach has been depicted in the Figure 15.

Saliency map of corresponding series of various emotions using method [10] and with our approach (Described in section 7) have been depicted in Figure 16.

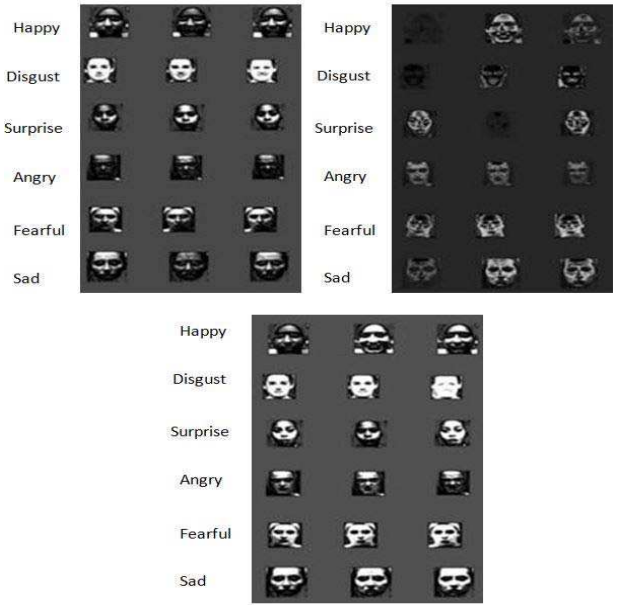

Figure 16: Top Left: Saliency Map of different level of various emotions (Happy, Disgust, Surprise, Angry, Fearful and Sad Series from Top to Bottom) using method [10], Top Right: Saliency Map of corresponding faces using proposed emotion scores, Bottom: Proposed Saliency Map after Fusing Frequency Tuned Technique [10] with Emotion score

In figure 16, in the proposed saliency map, salient face can be easily perceived because of higher intensity level for happy, disgust, surprise and angry emotion. But, it is quite difficult to finding the best salient face due to series of sad and surprise emotions. Therefore, for providing the better visualization, heat map of the corresponding faces of the saliency map has been generated using [29], which is depicted in Figure 17, for the corresponding series of emotions. The heat map area indicates about the region of the image that grabs our visual attention.

In the gaze-based heat map, one can better visualize about the most salient face, by analyzing the range of the heat spreads in the facial area.

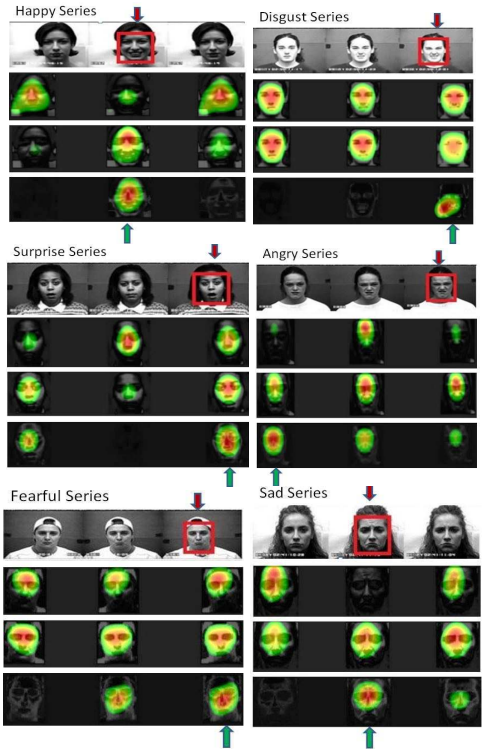

Figure 17: Corresponds to All Emotion Series: First Row: Ground truth indicating Most Salient Face indicated by Red Arrow Box, Second Row: Heat Map of Saliency Map generated by Method [10], Third Row: Heat Map of Saliency Map generated by Proposed Technique, Fourth Row: Heat Map of Saliency Map based on emotion saliency score by Proposed Technique with Most Salient Face indicated by Green Arrow

In Figure 17, we can easily identify, most of the results obtained through our proposed technique (i.e. Most Salient Face) are matching with the ground truth data. Ground truth results and the obtained results for most salient faces have been indicated through red and green arrow respectively.

In order to exhibit superiority of this method some more experimental results have been shown Figure18. We can easily observe that, in every emotion series, the faces with most expressive level of emotion are the most salient in the proposed map.

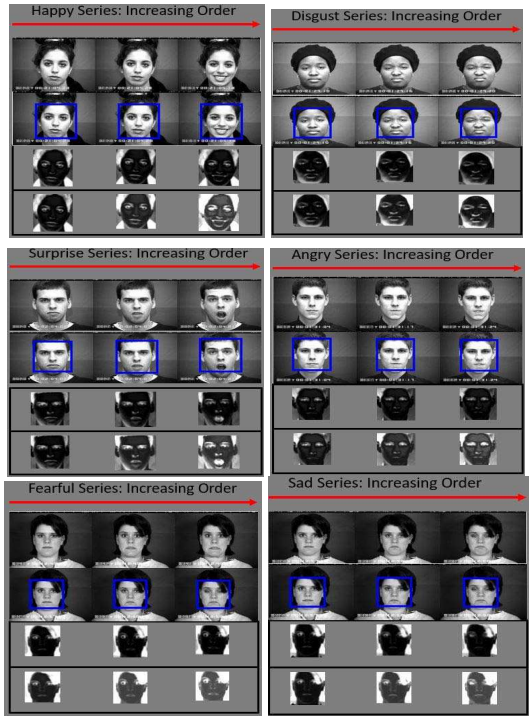

Figure 18: Corresponds to All Emotion Series: First Row: Input Image, Second Row: Detected Faces using [22], Third Row: Saliency Map generated by Method [10], Fourth Row: Saliency Map based on emotion saliency score by Proposed Technique

9. Conclusion

In this paper, we have proposed a new saliency model which can determine the most salient face in a series of faces, based on the low-level features and its associated emotion. Low level features have been extracted using frequency tuned saliency method [10]. Emotion Scores have been computed using deep CNN. Relative emotion saliencies of the faces have been computed using the emotion scores and the relative spatial distance among the faces. Now, these saliencies scores of the corresponding faces have been fused with method [10] for generating the final saliency map, which incorporates the emotion effect too. Moreover, Heat Map has been generated for determining the most salient regions. Our experimental results have been validated in two ways, statistically and perceptually, with the ground truth and it is found to be satisfactory. Therefore, this proposed technique can give new direction to the researchers for modeling the saliency for the things which exhibits emotions and expression. This work can be useful for developing the intelligent vision system for better visual security and surveillance.

References

[1] P. Ekman, Facial expression and emotion. American psychologist, 48(4):384, 1993

[2] B.L Fredrickson, The value of positive emotions: The emerging science of positive psychology is coming to understand why it's good to feel good. American scientist, 91(4):330-335, 2003

[3] K.N. Ochsner and J.J. Gross. The cognitive control of emotion. Trends in cognitive sciences, 9(5): 242-249, 2005

[4] R. Pal, R. Srivastava, S. K. Singh and K. K. Shukla. Computational models of visual attention: a survey. Recent Advances in Computer Vision and Image Processing: Methodologies and Applications (eds.):54-76, 2013.

[5] R. Pal, A. Mukherjee, P. Mitra and J. Mukherjee. Modelling visual saliency using degree centrality. IET Computer Vision, 4(3): 218-229, 2010.

[6] R. Mafrur, I. G. D. Nugraha and D. Choi. Modeling and discovering human behavior from smartphone sensing life-log data for identification purpose. Human-centric Computing and Information Sciences, 5(1): 31, 2015.

[7] L. M. Brown, A. W. Senior, Y.L. Tian, J. Connell, A. Hampapur, C. F. Shu, H. Merkl, and M. Lu. Performance evaluation of surveillance systems under varying conditions. In Proceedings of IEEE Pets Workshop: 1-8: January 2005.

[8] S. F. Taylor, K. L. Phan, J. C. Britton and I. Liberzon. Neural response to emotional salience in schizophrenia. Neuropsychopharmacology, 30(5): 984, 2005.

[9] J. Harel, C. Koch and P. Perona. Graph-based visual saliency. In Advances in neural information processing systems, 545-552, 2007.

[10] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In Computer vision and pattern recognition. IEEE cvpr 2009.1597-1604, June 2009.

[11] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price and R. Mech. Minimum barrier salient object detection at 80 fps. In Proceedings of the IEEE International Conference on Computer Vision: 1404-1412, 2015

[12] L. S. Goferman, Zelnik-Manor and A. Tal. Context-aware saliency detection. IEEE transactions on pattern analysis and machine intelligence, 34(10): 1915-1926, 2012

[13] M. M. Cheng, N. J. Mitra, X. Huang, P. H. Torr and S. M. Hu. Global contrast based salient region detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(3), 569-582, 2015.

[14] M. Jiang, J. Xu, and Q. Zhao. Saliency in crowd. In European Conference on Computer Vision, Springer, Cham. 17-32. September 2014,

[15] V. Mahadevan, W. Li, V. Bhalodia and N. Vasconcelos. Anomaly detection in crowded scenes. In Computer Vision and Pattern Recognition (CVPR), IEEE Conference, 1975-1981, June 2010.

[16] P. Vuilleumier and S. Schwartz. Emotional facial expressions capture attention. Neurology, 56(2): 153-158, 2001.

[17] J. Kucerova. Saliency map augmentation with facial detection. In Proceedings of the 15th Central European Seminar on Computer Graphics, 2011.

[18] R. K. Kumar, J. Garain, D. R. Kisku and G. Sanyal. A novel approach to enlighten the effect of neighbor faces during attending a face in the crowd. In TENCON, IEEE Region 10 Conference, 1-4, November 2015.

[19] R. K. Kumar, J. Garain, D. R. Kisku and G. Sanyal. Constraint Saliency-Based Intelligent Camera for Enhancing Viewers Attention towards Intended Face, Pattern Recognition Letters, Elsevier, https://doi.org/10.1016/j.patrec.2018.01.002, In Press, 2018.

[20] FERC 2013, Form 714 – Annual Electric Balancing Authority Area and Planning Area Report (Part 3 Schedule 2) 2006–2012 Form 714 Database, Federal Energy Regulatory Commission, 2013.

[21] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE Computer Society Conference, 94-101, June 2010.

[22] P. Viola and M. J. Jones. Robust real-time face detection. International journal of computer vision, 57(2):137-154, 2004.

[23] Lienhart, Rainer, and Jochen Maydt. An extended set of haar-like features for rapid object detection. Image Processing. IEEE International Conference, Vol. 1, 2002.

[24] Y. LeCun, LeNet-5, Convolutional neural networks. URL: http://yann. lecun. Com/exdb/lenet (2015).

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep con-volutional neural networks. In Advances in neural information processing systems, 1097-1105, 2012.

[26] S. Tokui, K. Oono, S. Hido, and J. Clayton. Chainer: a next-generation open source framework for deep learning. In Proceedings of workshop on machine learning systems (LearningSys) in the twenty-ninth annual conference on neural information processing systems (NIPS), Vol. 5, December 2015.

[27] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(11):1254–1259, 1998

[28] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. IEEE Conference on Computer Vision and Pattern Recognition, 2007.

[29] A. Garcia-Diaz, V. Leboran, X. R. Fdez-Vidal, and X.M. Pardo. On the relationship between optical variability, visual saliency, and eye fixations: A computational approach. Journal of vision, 12(6), 17-17, 2012.