International Journal of Software Engineering and Its Applications
Vol. 9, No. 7 (2015), pp. 85-92
http://dx.doi.org/10.14257/ijseia.2015.9.7.09

Estimating Normalized Attention of Viewers on Account of Relative Visual Saliency of Faces (NRVS)

Ravi kant kumar[1*], Jogendra Garain[2], Goutam Sanyal[3] and Dakshina Ranjan Kisku[4]

[1,2,3,4] *Department of Computer Science and Engineering*
National Institute of Technology Durgapur
Durgapur – 713209, West Bengal, India
{vit.ravikant, jogs.cse, nitgsanyal, drkisku}@gmail.com

Abstract

Humans psychological and behavioral understanding often lead to make natural decision which accurately identifies and remembers the faces which are highly appreciated or criticized by themselves in comparing to the normal viewed faces, in terms of beauty, ugliness or unique appearance. It happens due to human psychology of being biased towards the salient face in the process of face recognition and identification. This paper attempts a novel method to measure, how our attention is more restricted towards some particular faces in the crowd. This restricted attention is strongly guided by the relative visual saliency of these faces. In this paper, normalized relative visual saliency (NVRS) of the faces is evaluated using their intensity values modulated with respective spatial distance. Experiment has been carried out on test image dataset via bottom up approach. The experimental results are found to be encouraging and accuracy has also been measured exhibiting efficacy of the proposed approach.

Keywords: *Face Recognition, Visual Saliency, Face Identification, Visual Attention.*

1. Introduction

Human beings have the extraordinary ability to attend the different faces with their unique identity and continuously getting visual information from surrounding scenes. But, one can't process all the visual inputs that falls on the retina of the eye. The human brain intelligently selects only little visual information, which can reach the deeper levels of the brain. This action is known as selective attention [1, 2]. Selective visual attention is an imperative component of human vision system that quickly notices only relevant regions from the visual input. These regions draw attention due to being more salient with respect to the surrounding regions. Saliency, in the absence of any external stimuli, is determined by low level features like intensity, color, *etc.* [1-3]. Generally, visual saliency of an object is determined by the feature difference with the surrounding objects. The natural capability of the human vision system to view an image is biased towards some certain area in the image due to the relative saliency with respect to their surroundings.

Computer vision researchers get inspired from human's attention mechanisms to model it in the same way as human brain works. Some recognition systems have been developed based on visual saliency like, selective attention-based method for face recognition [4], Attention capture by faces [5], saliency map augmentation with facial detection [6], 3D face recognition using Kinect [7].

Rest of this paper is organized as follows. In Section 2, mathematical formulation for calculating the saliency score of the faces is discussed. The proposed approach is described briefly in Section 3. Experimental validation of the proposed technique is

ISSN: 1738-9984 IJSEIA

presented in Section 4. Last section draws the concluding statements and future work.

2. Evaluation and Measurement of Saliency

An object in a scene is more attentive due to its higher saliency value. Saliency of an object is determined by its feature dissimilarity (contrast) with respect to the surrounding objects. In [8], contrast is measured by centre surround difference of multi-scaled feature maps. In some other approaches [9-10], contrast of a region is evaluated by the joint effect of feature dissimilarity and positional proximity of the surrounding regions where both dissimilarity and nearness play major role for calculating saliency. There exist various vision models and most of them are built based on saliency [11-15]. The proposed work which is inspired from [10] makes use of mathematical formulation and measurement of saliency. In model [10], quad-tree based decomposition is applied to form homogeneous block. A block of homogeneous pixels signify one node in the graph. The edge-weight E_{ij} between any couple of nodes i and j is expressed by their feature difference, modulated by positional proximity.

$$E_{ij} = D_{f_{ij}} e^{-D_{S_{ij}}^2 / 2\sigma^2} \tag{1}$$

In Equation (1), $D_{f_{ij}}$ is the feature dissimilarity between the nodes i and j. It is computed as the absolute difference of mean feature values of the concerned blocks. In the proposed work, only gray-scale images have been considered. Therefore, the term 'feature' signifies intensity in this context. The spatial distance $D_{S_{ij}}$ between any blocks i and j is measured as the Cartesian distance between the center points of these blocks. As recommended in [10], the saliency of a block, represented by a node i, can be expressed as:

$$S_i = E_{ij} = \sum_j D_{f_{ij}} . e^{-D_{S_{ij}}^2 / 2\sigma^2} \tag{2}$$

Equation (1), for the same feature values of two nodes i and j, results their relative saliency value as zero. Hence, positional proximity between nodes i and j does not contribute any role for this particular case. Therefore Equation (1) is not appropriate to handle such cases. So this paper proposed a new mathematical model in order to beat this problem.

The proposed mathematical equations for calculating saliency are as below.

$$E_{ij} = (D_{f_{ij}} + 1/\sigma\sqrt{2\pi}) e^{-D_{S_{ij}}^2 / 2\sigma^2} \tag{3}$$

and, saliency is given by,

$$S_{ij} = \sum_j (D_{f_{ij}} + 1/\sigma\sqrt{2\pi}) . e^{-D_{S_{ij}}^2 / 2\sigma^2} \tag{4}$$

Where $(1/\sigma\sqrt{2\pi})$ and σ are the Gaussian coefficient and standard deviation respectively. Since, saliency of a face is affected inversely with spatial distance of surrounding faces.

For getting the attention of the faces on the basis of spatial distance modulated with the feature difference (contrast), the novel mathematical formulations for calculating edge weight and saliency are described as follows.

$$E_{ij} = \frac{1}{(D_{S_{ij}} + \sigma\sqrt{2\pi})} e^{-(\frac{1}{(D_{f_{ij}} + \sigma\sqrt{2\pi})^2})/2\sigma^2} \quad (5)$$

$$S_i = \sum E_{ij} = \sum_j \frac{1}{(D_{S_{ij}} + \sigma\sqrt{2\pi})} e^{-(\frac{1}{(D_{f_{ij}} + \sigma\sqrt{2\pi})^2})/2\sigma^2} \quad (6)$$

In Equation (5) and (6), E_{ij} and S_i denote edge weight and saliency respectively. For getting the similar and dissimilar attention of the faces on the basis of feature difference (contrast), modulation with spatial distance and vice versa as per equation (4) and (6), the normalised saliency can be proposed as follows.

$$S_{nor} = \frac{\sum_j \frac{1}{(D_{S_{ij}} + \sigma\sqrt{2\pi})} e^{-(\frac{1}{(D_{f_{ij}} + \sigma\sqrt{2\pi})^2})/2\sigma^2}}{\sum_j (D_{f_{ij}} + 1/\sigma\sqrt{2\pi}).e^{-D_{S_{ij}}^2/2\sigma^2}} \quad (7)$$

3. Proposed Approach

The proposed approach to get saliency values of all the faces in the input image and to determine the most salient face, involves following steps:

Algorithm 1: Calculating Saliency Value

Input: Image having set of faces
Output: Salient Face

1. For i= 1 to N (No of faces in the Input Image) do
   Find Center coordinates of the faces, C(Center Vector) = [C_1,C_2,C_3,C_4, ……C_n]
2. Extract all the faces *(N)* from the input Image
3. Find 'n_i' Homogeneous regions (Appropriate No of Segments) after applying Mean Shift Segmentation Algorithm [16].
4. For j= *1 to* n_i (For all the segments *'n_i'* of each face)
   Identify the Intensity (I_j), for all 'j' segments of each *'i'* face.
5. Find Average Intensity of each face as below:

$$\text{Avg}(I_i) = \frac{\sum_j I_j}{n_j}$$

6. Store these intensity values of all faces in the intensity vector (I) = [I_1, I_2, I_3, I_4 … I_N].
7. Calculate Saliency value (S_1) for each faces of input image based on Center vector (C) and Intensity vector (I), using equation (4).
8. Calculate Saliency value (S_2) for each faces of input image based on Center vector (C) and Intensity vector (I), using equation (6).
9. Obtained Normalized Saliency values (S_{Nor}) by using equation (7).
10. Gets the face having highest normalized saliency value (S_{Nor}), as the most attentive face.

The algorithm for the proposed approach is as below:

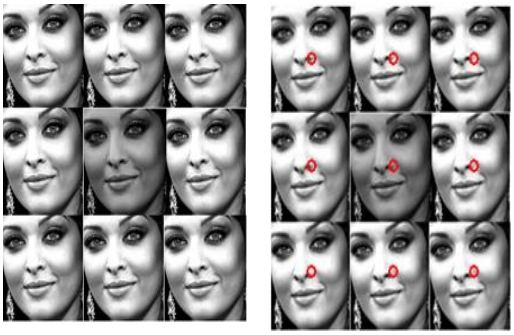
The above steps can be applied to the all types of images like color, black and white, gray scale *etc*. For color image saliency will be calculated based on RGB values difference and the spatial distance among the faces. In this paper, experiment has been done on Gray-scale images. Therefore, intensity is considered as the key feature for calculating saliency. Other low level features like orientation and texture

insignificantly contributing to saliency have been ignored, for simplicity. A face may also more salient due to its shape size, structure, expression, pose *etc*., but in this experiment, the main focus is to estimate the saliency of faces based on variation of feature difference and the spatial distance among the faces. Therefore, to ignore the effects of structure, size and expression variations on saliency and make much fair with the saliency estimation only due to low level features (here intensity), same faces with different intensity and spatial location is incorporated as an input image.

The overall proposed algorithm is described as below:

As input image consists of same dimension of faces, so their centers (C) can be found by using simple geometry. Center coordinates of all the face of input images (Row wise: Left to Right) in Figure 1(b), are depicted in Table 1.

Table 1. Center Coordinates of Faces



1.(a) 1.(b)

Face No	X-Coordinate	Y-Coordinate
1	36.83	36.16
2	110.50	36.16
3	184.16	36.16
4	36.83	108.50
5	110.50	108.50
6	184.16	108.50
7	36.83	180.83
8	110.50	180.83
9	184.16	180.83

Figure 1. (a) Input Image, (b) Center of Faces

All the faces are extracted and segmented by applying mean shift algorithm [16], in order to make the homogeneous regions or blocks. A block is called homogeneous if all the belonging pixels to it have the same feature values. Here, in this proposal, gray scale images are taken for the experiment so, intensity is the main feature for making the homogeneous block. Based on dissimilarity, low contrast (centre-middle) and high contrast faces (centre-surrounding) exist in the input image (Figure 1(a)).

2.(a) 2.(b)

Figure 2. (a) Extracted Faces and Formation of Homogeneous Block (Low and High Contrast), (b) Corresponding Segmented Faces of Input Image

Mean intensity value of each faces is calculated by the following formula:

$$Mean(I) = No\ of\ Pix_{Seg(i)} * Intensity_{Seg(i)}) \Big/ Total\ Pixels \quad (8)$$

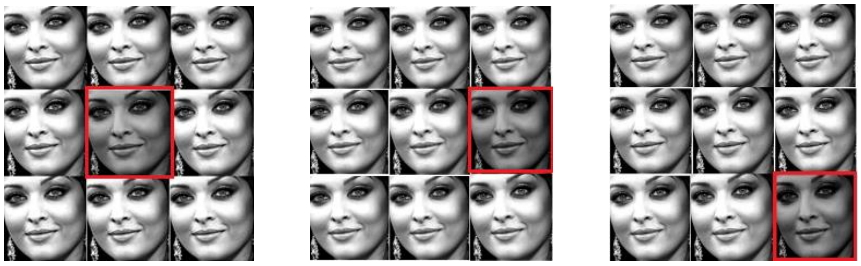
where *Seg (i)* represents i^{th} segment and *Mean (I)* is the average intensity of corresponding segment. Obtained intensity of all the faces of input image are described in Table 2.

Table 2. Obtained Intensity Values of Faces (Row wise: Left to Right)

Face No	1	2	3	4	5	6	7	8	9
Avg. Intensity	147	147	147	147	104	147	147	147	147

Now, Cartesian (Spatial) distances among all the face centers are calculated. Finally, saliency values of all the faces are obtained by using Equation (4) and (6). For find out similar and dissimilar attentive faces, normalized saliency is calculated by using Equation (7).

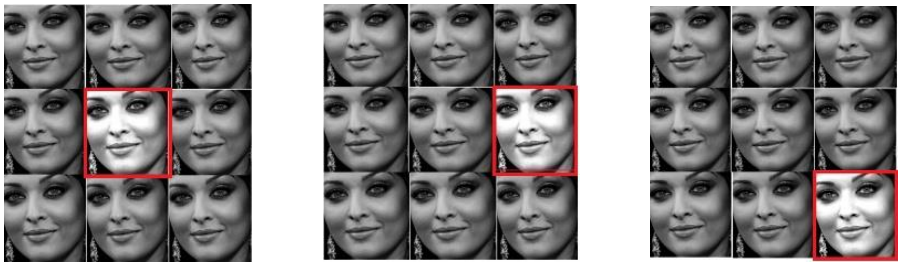
In this proposal, saliency is calculated not only on the basis of simply contrast difference, but also with the effect of spatial distance of the surrounding faces. In the following figure (Figure 4(a), 4(b) and 4(c)), the most attentive face is shown (Highlighted in Red).



3.(a) 3.(b) 3.(c)

Figure 3. (a) Salient Face at Center (b) Salient Face at Last of Middle Row (c) Salient face at Last of Last Row

Similarly, the most salient face (Highlighted in Red) are shown in Figure 4(a), 4(b) and 4(c).



4.(a) 4.(b) 4.(c)

Figure 4. (a) Salient Face at Center (b) Salient Face at Last of Middle Row (c) Salient Face at Last of Last Row

4. Experiment Results and Validation

Experiment has been conducted for 55 different sets of gray scale images each containing 6 to 12 faces. The face having highest saliency score indicates the most attentive face in that set. The attentiveness order of the corresponding faces found to be same as per saliency obtained from mathematical modeling, Equation (4) and Equation (6). Similar attentive faces as well as the dissimilar face (most salient face) have also been found by calculating the normalized saliency using Equation (7). In Table 3, Table 4 and Table 5 the normalized score of similar attentive faces

the saliency is ranging in between (0.0270-0.0277), (0.0242-0.0310) and (0.0219-0.0278) respectively. Whereas in the respective table the saliency values of most salient faces are 0.0523, 0.0551 and 0.0585 respectively, which is being far ahead from the saliency, range of similar faces. In various previous researches it has been proved that, more salient object provides more visual perception and attention. Therefore, it also works in the same way with faces in crowd flux. This proposed experiment gives satisfactory result with 95.32% of accuracy.

The saliency scores of Figure 3{(a),(b),(c)} and Figure 4{(a),(b),(c)}, obtained by equation (4) and (6), say saliency (*S1*) and saliency (*S2*) respectively, are shown in Table 3, Table 4 and Table 5. The saliency values of the subsequent faces of Figure {3(a), 4(a)}, {3(b), 4(b)} and {3(c), 4(c)} are same due to the same effect of intensity difference and proximity with their surrounding faces. The obtained normalized saliency scores using equation (7), is also given in the last column of the tables.

Table 3. Saliency Scores of Figure 3(a) OR 4(a)

Face No	Saliency (S_1)	Saliency (S_2)	Normalised Saliency
1	0.6433	0.0174	0.0270
2	0.8069	0.0224	0.0277
3	0.6433	0.0174	0.0270
4	0.8024	0.0222	0.0276
5	1.7635	0.0923	0.0523
6	0.8024	0.0222	0.0276
7	0.6433	0.0174	0.0270
8	0.8069	0.0224	0.0277
9	0.6433	0.0174	0.0270

Table 4. Saliency Scores of Figure 3(b) OR 4(b)

Face No	Saliency (S_1)	Saliency (S_2)	Normalised Saliency
1	0.5888	0.0143	0.0242
2	0.7773	0.0189	0.0243
3	0.6729	0.0209	0.0310
4	0.7327	0.0165	0.0225
5	0.9616	0.0241	0.0250
6	1.4256	0.0786	0.0551
7	0.5888	0.0143	0.0242
8	0.7773	0.0189	0.0243
9	0.6729	0.0209	0.0310

Table 5. Saliency Scores of Figure 3(c) OR 4(c)

Face No	Saliency(S_1)	Saliency(S_2)	Normalised Saliency
1	0. 5640	0.0133	0.0235
2	0. 7241	0.0159	0.0219
3	0. 6043	0.0150	0.0248
4	0. 7195	0.0159	0.0220
5	0. 9333	0.0208	0.0222
6	0. 8036	0.0224	0.0278
7	0. 6020	0.0149	0.0247
8	0. 8057	0.0222	0.0275
9	1. 1484	0.0672	0.0585

5. Conclusion and Future Work

This paper recommends an innovative approach that explores and implements the attentiveness of a salient face in the crowd. Gray scale image is considered for the experiment. Hence, intensity is taken as the fundamental feature. For making the experiment simple and more relevant to the gray images, performance is measured with intensity, other parameters like color, orientation, texture etc. have been overlooked, in this work. The newness of this work lies in the two aspects of mathematical modeling for calculating saliency, based on feature differences in modulation with the spatial distances and vice versa. Subsequently, normalized saliency is also evaluated for obtaining the similar and dissimilar (most salient) attentive faces in the crowd. Saliency score

determines the attentiveness of the faces. Normalized saliency scores majors the similarity and dissimilarity among the faces. Accuracy of the obtained results has also been measured and it found as satisfactory. Considering RGB values as the feature differences, future work can be extend to color images.

References

[1] L. Itti and C. Koch, "Computational Modeling of Visual Attention", Nature Reviews Neuroscience, vol. 2, no. 3, **(2001)**, pp. 194-203.
[2] R. Pal, "Computational Models of Visual Attention: A Survey", Recent Advances in Computer Vision and Image Processing: Methodologies and Applications, (R Srivastava, S. K. Singh, and K. K. Shukla), IGI Global, **(2013)**, pp. 54-76.
[3] J. M. Wolfe and T. S. Horowitz, "What Attributes Guide the Deployment of Visual Attention and how do they do it?", Nature Reviews Neuroscience, vol. 5, **(2004)**, pp. 495-501.
[4] A. A. Salah, E. Alpaydın and L. Akarun, "A Selective Attention-based Method for Visual Pattern Recognition with Application to Handwritten Digit Recognition and Face Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 3, **(2002)**, pp. 420-425.
[5] S. R. Langton, A. S. Law, A. M. Burton and S. R. Schweinberger, "Attention Capture by Faces, Cognition", vol. 107, no. 1, **(2008)**, pp. 330-342.
[6] J. Kucerova, "Saliency Map Augmentation with Facial Detection", Proceedings of the 15th Central European Seminar on Computer Graphics, **(2011)**.
[7] P. Yang, "Facial Expression Recognition and Expression Intensity Estimation", PhD Thesis, New Brunswick Rutgers, The State University of New Jersey, **(2011)**.
[8] L. Itti, C. Koch and E. Niebur, "A Model for Saliency based Visual Attention for Rapid Scene Analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 11, **(1998)**, pp. 1254-1259.
[9] J. Harel, C. Koch and P. Perona, "Graph-based Visual Saliency", Advances in Neural Information Processing Systems, **(2006)**, pp. 545–552.
[10] R. Pal, A. Mukherjee, P. Mitra and J. Mukherjee, "Modelling Visual Saliency using Degree Centrality", IET Computer Vision, vol. 4, **(2010)**, pp. 218-229.
[11] Y. F. Ma, L. Lu, H. J. Zhang and M. Li, "A User Attention Model for Video Summarization", Proceedings of the 10th ACM International Conference on Multimedia, **(2002)**, pp. 533-542.
[12] M. Cerf, J. Harel, W. Einhäuser and C. Koch, "Predicting Human Gaze using Low-level Saliency Combined with Face Detection", Advances in Neural Information Processing Systems, **(2008)**, pp. 241-248.
[13] D. Neumann, M. L. Spezio, J. Piven and R. Adolphs, "Looking You in the Mouth: Abnormal Gaze in Autism Resulting from Impaired Top-down Modulation of Visual Attention", Social Cognitive and Affective Neuroscience, vol. 1, no. 3, **(2006)**, pp. 194-202.
[14] P. Vuilleumier, J. L. Armony, J. Driver and R. J. Dolan, "Effects of Attention and Emotion on Face Processing in the Human Brain: An Event-related fMRI Study", Neuron, vol. 30, no. 3, **(2001)**, pp. 829-841.
[15] C. Breazeal and B. Scassellati, "A Context-dependent Attention System for A Social Robot", Proceedings of the 16th International Joint Conference on Artificial Intelligence, **(1999)**.
[16] D. Comaniciu and P. Meer, "Mean shift: A Robust Approach towards Feature-Space Analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, **(2002)**, pp. 603-619.

Authors

Ravi Kant Kumar, at present, he is pursuing Ph.D. from NIT Durgapur (India) in the CSE department. In 2014, he completed his M.Tech, in Information Technology from Central University of Hyderabad, India. Prior to this he worked as a Software Developer. His areas of interest are computer vision, pattern recognition image processing, fuzzy logic and mathematical modeling.

Jogendra Garain, he is Pursuing his PhD from NIT Durgapur (India) in the department of CSE. He completed his Master of Technology from University of Calcutta and B.Tech from HIT in the department of CSE. He has 6 years of teaching experience as an Assistant Professor in DIATM, Durgapur. His area of interest is image processing, computer vision, pattern recognition and theory of computation.

Prof. Goutam Sanyal, is designated as Professor in the Department of Computer Science and Engineering and Dean (FW) of NIT Durgapur, India. His qualifications are Bachelors of Engineering (B.E), Masters of technology (M.Tech), PhD (Engineering), FIE (India), MIEEE. Dr. Sanyal has more than 150 research papers in the reputed international journals and conferences. He has a work experience of 28 years in teaching and research along with PhD Guidance. His areas of interests are Computer Architecture, Computer Graphics, Computer Vision, Image Processing, VLSI, and Mathematical Modeling.

Dr. Dakshina Ranjan Kisku, is designated as an Assistant Professor in the Department of CSE, NIT Durgapur. His qualifications are B.E, M.E and PhD (Engineering) from Jadavpur University. He has about 12 years of teaching experience. Dr. D.R. Kisku has more than 38 research papers in the reputed international journals and conferences. His areas of research are Biometric, Computer Vision, Pattern Classification, Affective Computing and Data Compression.