Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 54 (2015) 756 – 763
Procedia Computer Science

Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015)

Analysis of Attention Identification and Recognition of Faces through Segmentation and Relative Visual Saliency (SRVS)

Ravi Kant Kumar*, Jogendra Garain, Goutam Sanyal and Dakshina Ranjan Kisku

Computer Science and Engineering Department, National Institute of Technology, Durgapur, India

Abstract

Attentiveness, identification and recognition of a human face in a crowd play a starring role in the perspective of visual surveillance. Human vision system is not attending, identifying and recognizing all the faces with the same perception. It is biased towards some faces in the crowd due to their higher relative visual saliency and segment wise perception with respect to the surroundings faces. Using different computer vision-based techniques enormous researches have been carried out on attention, recognition and identification of the human face in context of different applications. This paper proposes a novel technique to explain and analyse how the attention, identification and recognition of a face in the crowd guided through segmentation and their relative visual saliency. The proposed method is stretched out the solution, using the concept of segmentation and relative visual saliency which is evaluated on the intensity values and respective spatial distance of the faces.

Peer-review under responsibility of organizing committee of the Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015)

Keywords: Face recognition; Identification; Relative visual saliency; Visual attention.

1. Motivation

The inspiration for writing this proposal is to provide an idea towards the development of the computer's vision as like human vision system. Humans, in reality, processing around 60 images per second with large no of pixels in each image. In reality, deeper level of brain has some limited capacity to processes all these visual information. So, in the beginning at lower lever of brain, these images are segmented into various meaningful regions and only the important or salient regions of the scene goes to the deeper level of brain, where attention, recognition and identification takes place[1,2]. Finally, the aim is to imitate the computer's vision similar to human vision but this goal is still a very long way away.

2. Introduction

Humans are expert at efficiently gathering information from their dynamic surrounding. Even they have huge dealing capability, human brain is unable to process all incoming visual information at a time. It logically filters out the bulk of incoming sensual information. For further processing, only filtered information goes to the deeper levels of

*Corresponding author.
E-mail address: vit.ravikant@gmail.com

Peer-review under responsibility of organizing committee of the Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015)
doi:10.1016/j.procs.2015.06.089

the brain. This mechanism is named as selective attention[1,2]. The filtered information belongs to the salient portions of the image or scene. Further, at inner brain, recognition of those attended objects takes place[2,3].

Any object (in this paper, face is considered as an object) is more salient, due to its dominating perceptual properties[1,2] like colour, intensity, orientation, shape, size, views etc. In the crowd, human vision system is not visiting, identifying and recognizing all the faces with uniform perception. In order to understand, the problem of machine recognition of human faces and its biased attentiveness towards the salient faces, enormous researches[4–11] has been done.

The idea behind developing the face recognition system is for person identification, verification and authorisation[12,13] without his knowledge or without giving extra duty of physical interaction to the any biometric system. Computer vision researchers have also derived several face recognition approaches[12–14]. Some of them are, Geometrical method to identifying the eyes and mouth in real scene[15], The Statistical shape model[16], Holistic matching methods based on PCA[17]. In Feature-based (structural) matching methods, the Hidden Markov model[18], Convolution Neural Network[19]. Some recognition systems have been developed based on visual saliency like, Selective attention-based method for face recognition[5], Attention capture by faces[6], Saliency map augmentation with facial detection[7], 3D face recognition using Kinect[9] and Visual perception based on eye movements[4] etc.

This paper proposes a novel method to illustrate how face attention, identification and recognition strongly determined by visual saliency and the appropriate segmentation. Features like intensity, color, orientation, etc. play a big role in determining saliency[20,21]. Moreover, significant feature-dissimilarity with other regions (mainly in the neighborhood) causes a region to pop-out as salient and to attract our gaze[20,21]. Experiment has been performed on gray scale images, so, intensity is incorporated as the main feature to calculate saliency score and segmentation. For making the experiment simple other features have been discounted. Since this experiment is functioning acceptably with one parameter (i.e. intensity), it can be employed that result will enhanced by including other relevant features.

Rest of this paper is organized as follows: In Section 3, Mathematical formulation for calculating the saliency score of the faces is discussed. The proposed approach is described briefly in Section 4. Experimental validation of the proposed technique is presented in Section 5. Finally, Section 6 draws the concluding statements and future work.

3. Proposed Mathematical Formulation to Evaluate Saliency

Significant feature dissimilarity (contrast) of a region draws our attention. In[21], center-surround difference is applied for multi-scaled image to obtaining the contrast. Here, the feature dissimilarity with respect to its surrounding is determined by the center surround difference. In some other models[22,23], contrast of a region is evaluated based on the feature dissimilarity with all the other regions of the image. However, in these approaches, feature difference is modulated with positional proximity. As a result, the dissimilarity with nearby regions contributes more in the saliency. This paper adopts a graph based model[23], in which both 'feature difference' and 'positional proximity' is considered for computing saliency. In model[23], quad-tree based decomposition is applied to form homogeneous block. A block of homogeneous pixels represent one node in the graph. The edge-weight E_{ij} between any pair of nodes i and j is expressed as feature difference (modulated by positional proximity) between the concerned block.

$$E_{ij} = |f_i - f_j| e^{-D^2_{S_{ij}}/2\sigma^2} \tag{1}$$

where f_i and f_j are the feature values of the blocks represented by nodes i and j. This is computed as the absolute difference between mean feature values of the concerned blocks. In this 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 nothing but the Cartesian distance between the centre points of these blocks. As recommended in[23], the saliencies of the nodes (or blocks) are determined by their degree which is confined by the summation of feature differences (modulated by positional proximity) with all other blocks. Therefore, the saliency of a block, represented by a node i, can be expressed as:

$$S_i = \sum_j E_{ij} = \sum_j |f_i - f_j| \cdot e^{-D^2_{S_{ij}}/2\sigma^2} \tag{2}$$

According to equation (1), for the same feature values of two nodes i and j, their relative saliency value will be zero. It means positional proximity between nodes i and j does not contributes any role for this particular case. Therefore, equation (1) is not appropriate to handle such cases. So, to include the contribution of positional proximity for such case the new equation for calculating relative saliency is proposed as below.

The proposed mathematical equation for calculating saliency is as below.

$$S_i = \sum_j (|f_i - f_j| + 1/\sigma\sqrt{2\pi}) \cdot e^{-D^2_{S_{ij}}/2\sigma^2} \tag{3}$$

where, $(1/\sigma\sqrt{2\pi})$ and σ are the Gaussian coefficient and standard deviation respectively. Equation (3), is capable to overcome the problem in the case of same feature value of node i and j.

4. Proposed Approach

The proposed approach to determine attentiveness of faces, their identification and recognition, involves following steps:

- Obtaining centre coordinates of all the faces of the input image.
- Extracting all the individual faces from the input image and segment it into homogeneous regions.
- Finding the mean intensity of all the belonging segments of the corresponding segmented face.
- Identify these mean intensity values to the corresponding faces of the input image.
- Calculating the saliency values of all the faces of the input image by using Equation (3).
- Comparing the saliency values of the faces and getting the most salient face.
- Interpretation of attentiveness, identification and recognition of faces based on no of segments (obtained on same bandwidth and same face size) and saliency score.

The above approach can be experiment with the other low level features like colour, orientation, texture etc. In case of colour image relative visual saliency will be calculated by RGB values differences, modulated with the spatial distance of the faces. In this paper, experiment has been performed on gray-scale images. Therefore, for calculating saliency, intensity is taken as the leading feature. Other features like orientation, colour and texture have been ignored, as they are not much effective with the gray image. A face can be relatively more salient due to its shape, size, expression, pose etc. In this research, the main focus is to estimate the saliency of faces based on feature difference, modulated with their spatial distance. Therefore, to justify the relative saliency only with low level features (here intensity) of the faces, the other external effects like structure, size and expression variations, has been ignored. For making the experiment more suitable and saving the saliency distraction due to background has also been resolved by taking the input image without background.

The details of the proposed steps are described in the following subsections.

4.1 Obtaining centers of all the faces

In this paper all the faces belonging to input image are of same size without any background, therefore, their center coordinates is obtained by using simple geometry. Center (C) of the input image is found as $C(r/2, c/2)$, where r = no of rows, c = no of columns of the input image. Obtained center coordinates for all the faces of input images (Row wise: Left to Right) in Fig. 1, are described in Table 1.

4.2 Face extraction and formation of homogeneous region

Unlike quad-tree based decomposition for the formation of rectangular homogeneous block[23], this paper adopts mean shift based image segmentation technique[24] by adjusting the various parameters like special, color bandwidth and no of regions to find out a set of homogeneous regions. Segmentation with considering more numbers of segments enhances the recognition accuracy because more no of segments will participate for calculating visual saliency for the recognition, but it makes the implementation more complex and also takes more also takes more processing time

Table 1. Center points of faces (row wise).

Face No	X-Coordinate	Y-Coordinate
1	34.33	41.83
2	103.00	41.83
3	171.66	41.83
4	34.33	125.50
5	103.00	125.50
6	171.66	125.50
7	34.33	171.66
8	103.00	171.66
9	171.66	171.66

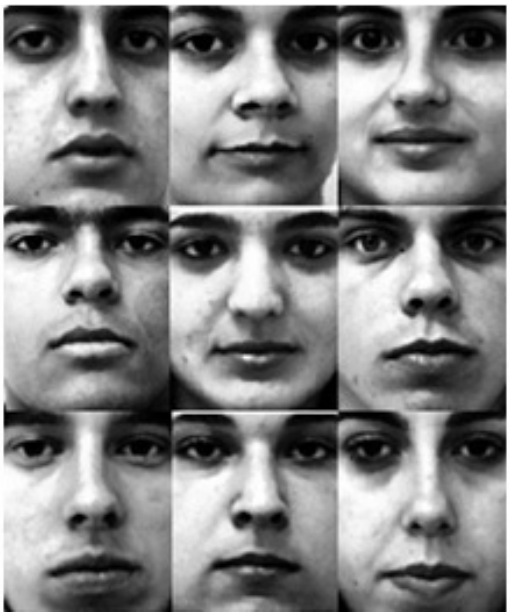

Fig. 1. Input image.

and cost. So for balancing these issues it is better to make an appropriate (not very less or more) no of segment by adjusting special, color bandwidth and no of regions during the segmentation. In this paper experiment has been done on gray scale images, so, intensity difference is the prominent factor of the formation of the segments in the form of homogeneous blocks or nodes. A homogeneous block indicates that all the belonging pixels of a particular segment have the same intensity values. In some cases, it may vary more or less due to inappropriate adjustment of the spatial and color bandwidth with respect to the no of segment.

4.3 Finding mean intensity of the faces of the input image

Every pixels belonging to the same segment have more or less same intensity value. The intensity value of a particular segment is determined by the mean intensity of all its belonging pixels. In order to find the central feature value of a particular segmented face image, average intensity value with respect to all the segments and the their belonging pixels is calculated. Mean intensity value is calculated by equation (4) as below:

$$\text{Mean}(I) = (\text{No of Pix}_{\text{Seg}(i)} * \text{Intensity}_{\text{Seg}(i)})/\text{Total Pixels} \tag{4}$$

where Seg (i) and Mean (I) represents i^{th} segment and their mean intensity of the segments.

Table 2. No of segments of faces (row wise).

Face number	Spatial bandwidth	Colour bandwidth	No of segments (Obtained)
1	10	7	12
2	10	7	12
3	10	7	12
4	10	7	9
5	10	7	12
6	10	7	11
7	10	7	13
8	10	7	11
9	10	7	10

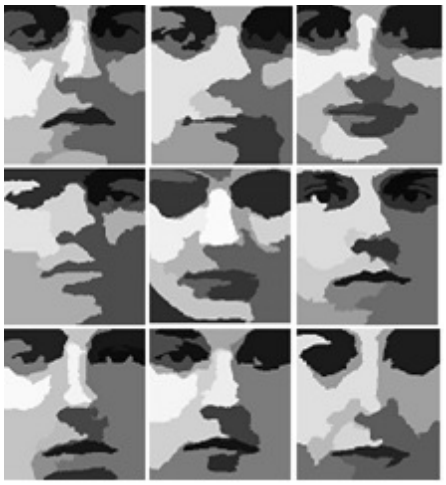

Fig. 2. Segmented image of faces.

Table 3. Obtained intensity values of faces (row wise: left to right in Fig. 2).

Face No	1	2	3	4	5	6	7	8	9
Avg. Intensity	125	120	132	122	134	110	122	123	135

Obtained intensity values for all the faces of input images (Row wise: Left to Right) of Fig. 2, are given in Table 3.

4.4 Calculating the saliency values of faces of the input image

Once the intensity values and the centre point of each faces of an image is known, intensity differences and spatial distances among all the faces are calculated. Spatial distance between two faces is the Cartesian distance from their centres. Finally, saliency values of all the faces are calculated by using Equation (3) and (4). The obtained saliency values for all the faces of input images (Row wise: Left to Right) of Fig. 1, are described in Table 4.

4.5 Obtaining salient face

Saliency score of the face 5 (Highlighted in Red) is greater than rest all the faces of the input image as shown in Table 4. This indicates that face 5, in the input image draws extra attention and hence giving more attraction to the viewers. The relative saliency score of all the faces in the input image is shown in Table 4.

Table 4. Obtained saliency values of faces (Fig. 3 row wise: left to right).

Face number	Relative saliency
1	0.6256
2	0.8420
3	0.6920
4	0.7471
5	1.0697
6	0.9672
7	0.6137
8	0.8167
9	0.7189

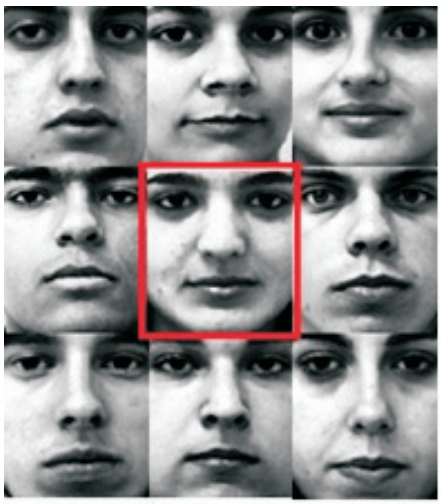

Fig. 3. Most salient face (highlighted in red).

4.6 Interpretation of attentiveness, identification and recognition of faces based on no of segments (obtained on same bandwidth and same face size) and saliency score

Attentiveness of an object in an image is determined by visual saliency[1,2,20,21]. So, here, more salient face gives more attraction. After attending any face, recognition takes place. Segmentation is the pre-process of the face recognition. It divides the images (here face image) into several meaningful homogeneous regions (segments). Area of interest for recognition increases with more no of segments. Therefore, face segmented with more no of regions (segments) gives better recognition[25]. Identification is the post-process of recognition (specific recognition). Proper segmentation along with more no of segments gives better identification.

After deep analysis, several experiments on various faces and obtained results, the combined effect of segmentation and relative visual saliency (RVS) is recommended in this paper as follows:

5. Experiment Validation

Experiment has been performed with 50 different sets of gray scale images having 6 to 12 faces in each set. The face having the highest saliency score results as most attentive face in that particular set. In various literatures survey, it has been proved that object having higher saliency score, provides more visual perception and attention. Similarly, in many previous researches it has been proved that proper segmentation and appropriate no of segments gives better recognition and identification[25]. So in this proposed experiment, obtained relative saliency values with proper and

Table 5. Analysis of attention/recognition/identification.

No of segments	Saliency value	Effect on attention/recognition/identification
More	More	• More Attentive • Better Recognition • Better Identification (With Appropriate Segmentation)
More	Less	• Less Attentive • Average Recognition • Average Identification
Less	More	• More Attentive • Less Recognition • Less Identification
Less	Less	• Less Attentive • Less Recognition • Less Identification

appropriate segmentation, justifies it in the context of attention of faces. The summery of the combined effect of appropriate segmentation and relative visual saliency (RVS) of faces is inscribed in Table 5. Ground truth has been prepared based on perception of about 200 people with the experimental input images. The image with highest no of similar perception (about 89%) is considered as real ground truth. With this Ground truth the proposed experiment achieved 96.42% of accuracy.

6. Conclusion and Future work

This paper projects a novel approach that explores how segmentation and relative visual saliency determines the attention, recognition and identification of faces in the crowd. Experiment has been performed with gray-scale images. Hence, intensity is taken as the fundamental feature. For making the experiment simpler and to measure the performance with intensity, other features like colour, orientation, texture etc. have been ignored, in this work. The newness of this work lies in the mathematical formulation for calculating saliency based on feature differences in modulation with the spatial distances amongst the belonging faces in the crowd. The obtained saliency score determines the attentiveness of the faces. The experiment also measures the effect of appropriate segmentation on the recognition and identification. Finally, combined effect of segmentation and relative saliency has been proposed in the summary Table 5. Accuracy of the obtained results has been measured with the ground truth and it found as satisfactory. Considering RGB values as the feature differences, future work can be extend to colour images.

References

[1] L. Itti and C. Koch, Computational Modeling of Visual Attention, *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, March (2001).
[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*, pp. 54–76, September (2013).
[3] S. Frintrop, E. Rome and H. I. Christensen, Computational Visual Attention Systems and their Cognitive Foundations: A Survey, *ACM Transactions on Applied Perceptions*, vol. 7, (2010).
[4] D. Noton and L. Stark, Eye Movements and Visual Perception, *Scientific Ann*, pp. 34–43, (1971).
[5] 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 Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, issue 3, pp. 420–425, (2002).
[6] S. R. Langton, A. S. Law, A. M. Burton and S. R. Schweinberger, Attention Capture by Faces, *Cognition*, vol. 107, no. 1, pp. 330–342, (2008).
[7] J. Kucerova, Saliency Map Augmentation with Facial Detection, pp. 60–66, (2011).
[8] V. Goswami, S. Bharadwaj, M. Vatsa and R. Singh, RGB-D Face Recognition using Kinect, *Sixth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pp. 1–6, September (2013).
[9] P. Yang, Facial Expression Recognition and Expression Intensity Estimation, *PhD Thesis, New Brunswick Rutgers, The State University of New Jersey*, (2011).
[10] R. A. Khan, Detection of Emotions from Video in Non-Controlled Environment, *PhD Thesis, LIRIS-CNRS, University Claude Bernard Lyon 1, France*, (2013).
[11] 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.). *IGI Global*, pp. 54–76, September (2013).

[12] W. Zhao, R. Chellappa, P. J. Phillips and A. Rosenfeld, Face Recognition: A Literature Survey, *ACM Computing Surveys (CSUR)*, vol. 35, issue 4, pp. 399–458, (2003).
[13] A. S. Tolba, A. H. El-Baz and A. A. El-Harby, Face Recognition: A Literature Review, *International Journal of Signal Processing 2*, vol. 2, po. 88–103, (2006).
[14] Rabia Jafri, and Hamid R. Arabnia, A Survey of Face Recognition Techniques. *JIPS 5*, vol. 2, p. 41–68.
[15] Jun Wang, Lijun Yin and Jason Moore, Using Geometric Properties of Topographic Manifold to Detect and Track Eyes for Human-Computer Interaction, *ACM Transactions on Multimedia Computing, Communications and Applications (TOMCCAP) 3*, vol. 3, no. 4, (2007).
[16] T. F. Cootes, J. Christopher Taylor, David H. Cooper and J. Graham, Active Shape Models-their Training and Application, *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, (1995).
[17] Kirby, Michael and Lawrence Sirovich, Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces, *IEEE Transactions Pattern Analysis and Machine Intelligence*, no. 1, pp. 103–108, (1990).
[18] A. V. Nefian and M. H. Hayes III, Hidden Markov Models for Face Recognition, Choice 1, vol. 6, (1998).
[19] S. Lawrence, C. L. Giles, A. C. Tsoi and A. D. Back, Face Recognition: A Convolutional Neural-Network Approach, *IEEE Transactions on Neural Networks*, vol. 8(1), pp. 98–113, (1997).
[20] 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(6), pp. 495–501, June (2004).
[21] 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(11), pp. 1254–1259, November (1998).
[22] J. Harel, C. Koch and P. Perona, Graph-based Visual Saliency, *Advances in Neural Information Processing Systems*, pp. 545–552, (2006).
[23] R. Pal, A. Mukherjee, P. Mitra and J. Mukherjee, Modelling Visual Saliency using Degree Centrality, *IET Computer Vision*, vol. 4, pp. 218–229, September (2010).
[24] D. Comaniciu and P. Meer, Mean Shift: A Robust Approach towards Feature-Space Analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 603–619, May (2002).
[25] A. Rabinovich, A. Vedaldi and S. J. Belongie, Does Image Segmentation Improve Object Categorization? *Department of Computer Science and Engineering, University of California, San Diego*, (2007).