LEARNING GAUSSIAN MIXTURE MODEL FOR SALIENCY DETECTION ON FACE IMAGES

*Yun Ren**, *Mai Xu** †, *Ruihan Pan**, *and Zulin Wang**

∗ School of Electronic and Information Engineering, Beihang University, Beijing, 100191, China
† EDA Lab, Research Institute of Tsinghua University in Shenzhen, Shenzhen, China
ryoma123456@126.com, MaiXu@buaa.edu.cn, avrillavigne0610@126.com, and wzulin@buaa.edu.cn

ABSTRACT

The previous work has demonstrated that integrating top-down features in bottom-up saliency methods can improve the saliency prediction accuracy. Therefore, for face images, this paper proposes a saliency detection method based on Gaussian mixture model (GMM) , which learns the distribution of saliency over face regions as the top-down feature. Specifically, we verify that fixations tend to cluster around facial features, when viewing images with large faces. Thus, the GMM is learnt from fixations of eye tracking data, for establishing the distribution of saliency in faces. Then, in our method, the top-down feature upon the the learnt GMM is combined with the conventional bottom-up features (i.e., color, intensity, and orientation), for saliency detection. Finally, experimental results validate that our method is capable of improving the accuracy of saliency prediction for face images.

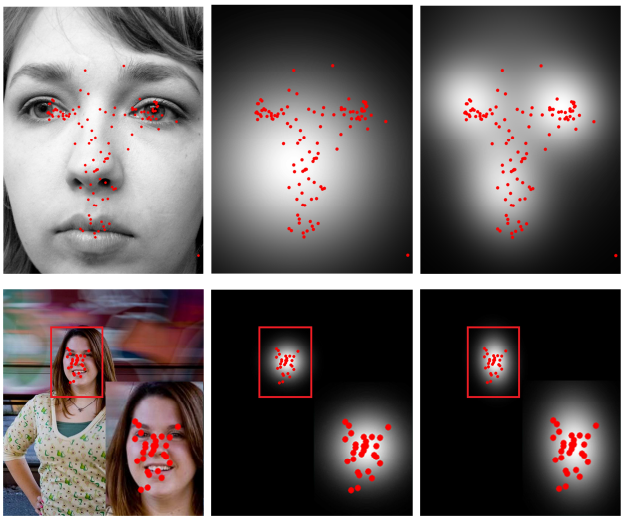
Index Terms— saliency detection, facial features, GMM

1. INTRODUCTION

Visual attention [1] has been widely studied in psychophysics, neurophysiology, and even computer vision societies [2]. With computation on features of either videos or images, saliency detection is an effective way to predict the visual attention attracted by different regions of a scene. As the output of saliency detection, the saliency map of an image or a video frame has been widely applied in object detection [3], object recognition [4], image retargeting [5], image quality assessment [6], and also image/video compression [7].

The existing methods on saliency detection can be classified into two categories: bottom-up and top-down methods. The representative bottom-up method on detecting image saliency is Itti's model [8], which combines center-surround features of color, intensity, and orientation together. Afterwards, Koch and Ullman [9] extended the Itti's model by incorporating the proto-object inference in the saliency map. Most recently, there has been extensive advanced work (e.g., [10–12]) on bottom-up saliency detection.

This work was supported by NSFC under grant number 61202139 and China 973 program under grant number 2013CB329006.



(a) Images with face (b) Isotropic GM (c) Learnt GMM

Fig. 1. Examples for saliency prediction vs fixations in face region, selected from [13]. The red dots represent the fixations recorded by the eye tracker. Note that both saliency and fixations belonging to face regions are displayed.

In fact, top-down visual features play an important role in determining the saliency of a scene. So, the top-down saliency detection methods have been extensively studied in [14–16]. Cerf *et al.* [15] found out that the face is an important top-down feature to attract human attention, as the faces were fixed on in 88.9% within first two fixations (7 subjects viewing 150 images with face) in their experiments. Therefore,they proposed to combine Viola & Jones (VJ) face detector [17] with Itti's model [8] for improving the saliency detection accuracy over face images. Since it is more reasonable to learn how important the top-down features are for attracting human attention, several state-of-the-art methods [13, 18, 19] have been proposed to apply machine learning algorithms in top-down saliency detection. For example, Zhao [19] utilized the fixations on face images to quantify the importance of face channel on attracting human attention.

Although the existing work has taken into account faces on saliency detection, it does not explore the actual distribution of eye fixations within faces. As shown in Figure 1, the simple assumption of isotropic Gaussian model (GM) for saliency in face regions [15, 19] has limitation on modeling

visual attention attracted by faces, especially for images with large scale[1] faces. In fact, saliency distribution, in the form of Gaussian mixture model (GMM), can be learnt from eye fixations on face images. Figure 1-(c) shows that the saliency with the learnt GMM distribution is more in accord with the ground-truth visual attention. This paper therefore proposes a GMM-based saliency detection method, which learns various GMMs across different face scales, for predicting visual attention in face images.

The contributions of this paper are listed as follows:

- We analyze the human visual attention on viewing images containing faces at different scales. The analytical results reveal that humans tend to be attracted by faces. Especially, when the scales of faces are large in the images, the majority of attention is drawn by facial features (e.g., eyes). Such results motivate our GMM-based method on saliency detection for face images.
- We model the human visual attention attended to face regions using GMM distribution. Specifically, we utilize Expectation Maximization (EM) algorithm [20] to learn GMMs for different face scales from the ground-truth fixations. Based on the learnt GMM , the distribution over face regions is incorporated as the top-down feature in saliency detection for face images.
- We establish a large database of 476 face images with 823 faces for the saliency analysis. These images with their corresponding fixations are selected from four state-of-the-art eye tracking databases. The database and the corresponding code are available on $www.ee.buaa.edu.cn/xumfiles/Attachment.html$.

2. MOTIVATION

In [15], as the top-down cue, face, is of great importance to draw human attention over face images. It is further intuitive that the facial features (e.g., eyes and mouth) may attract a large amount of human attention on large scale face. Thus, this section focus on figuring out how significant the face and facial features are for human attention. Section 2.1 discusses the eye tracking database we established for the statistical analysis. In Section 2.2, a method on automatically extracting the face and facial features is presented, as the preliminary for our statistical analysis of human visual attention. Section 2.3 shows the importance of face and facial features to human attention, via investigating the data of our eye tracking database.

2.1. Database

For saliency analysis of face region, we constructed a large database, which contains extensive face images with fixations viewed by several subjects. Specifically, the 476 images with 823 faces[2] were selected from four existing databases: the MIT database [13], FIFA database [21], PL database [22], and NUS database [23]. These images were selected with the following criteria:

- Each scale has sufficient numbers of images and fixations.
- Images with exaggerated expressions were not excluded in our database.
- Images with frontal (in which turning degree of head is less than $45°$) faces were included.

The images, eye tracking results, and the code for extracting fixations in our database are available on the web. This may have potential to facilitate the future work on the visual attention model of face regions.

2.2. Automatic detection on face and facial features

For analyzing the eye fixations on different parts of face, the region of face and its sub-regions for facial features have to be identified in an image. Generally speaking, our extraction technique is based on a real-time face alignment method [24]. To be more specific, several key feature points obeying the point distribution model (PDM) are located in an image using the method in [24], which combines the local detection (texture information) and global optimization (facial structure) together. Here, 68-point PDM is utilized to detect 68 key feature points in a face. Then, these points are connected to precisely extract the contours and regions of face and facial features.

2.3. Analysis on visual attention on face and facial features

Now, we concentrate on analyzing the visual attention on face and facial features. Here, we chose all images with one face from our database to avoid the influence among multiple faces on attracting attention in a scene. As a result, 281 images were finally selected with 36,497 fixations.

In order to analyze the quantified visual attention on the face, we plot in Figure 2 the average percentages of fixations (over the 281 images with 11,097 fixations on face and 25,400 fixations on the background) falling into face and background, respectively. We also plot in Figure 2 the proportions of pixels belonging to the face and background, respectively. Note that the faces were extracted using the method mentioned above. From this figure, we can see that although the faces averagely take up 3.7% of whole images, they attract 30.4% of eye fixations. This verifies that the visual attention on face is significantly more than that on background.

Beyond, there is an insight that the visual attention on face increases along with the enlarged size of face in the image. To validate such an insight, we show in Figure 3 the proportions of fixations on faces at different scales. Note that the scales of faces in those 281 images are clustered to 3 levels using the K-means algorithm.

Next, we move to the statistical analysis on the eye fixation points falling into different regions of face, to investigate the visual importance of facial features (i.e., left eye, right

[1] In this paper, scale means the proportion of pixel number belonging to face region in the image.

[2] There may exist more than one faces in an image.

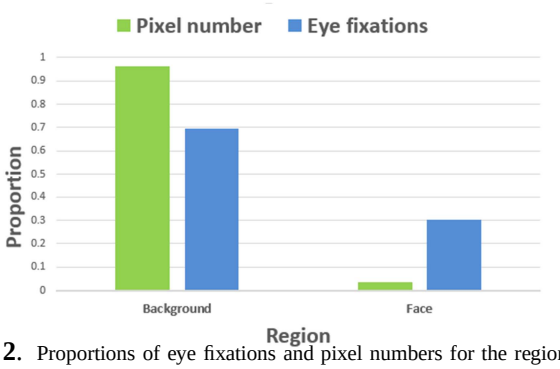


Fig. 2. Proportions of eye fixations and pixel numbers for the regions of face and backgroun

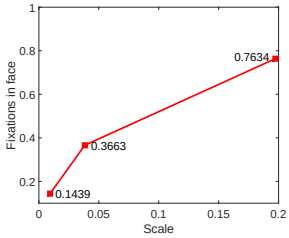


Fig. 3. Statistical results on proportions of eye fixations belonging to face at different scales. Note that the results are averaged over all 281 images.

eye, nose, and mouth). It is intuitive that the facial features are of great importance to visual attention, when the image is displayed with a close up of face. Therefore, it is interesting to find out the fixation numbers of facial features at various scales. Figure 4 shows the proportions of fixations belonging to right eye, left eye, nose, and mouth, averaged in all 281 images. From this figure, one may observe that more attention is paid to facial features, when the face has a large scale in the image. Figures 3 and 4 suggest that the visual importance of face and facial features enhances alongside the increased scale of face. It is worth mentioning that the attention enhancement of mouth is similar to that of face, whereas the attention on other facial features increases more sharply when approaching large scale. Thus, facial features need to be taken into consideration for saliency detection on large scale faces.

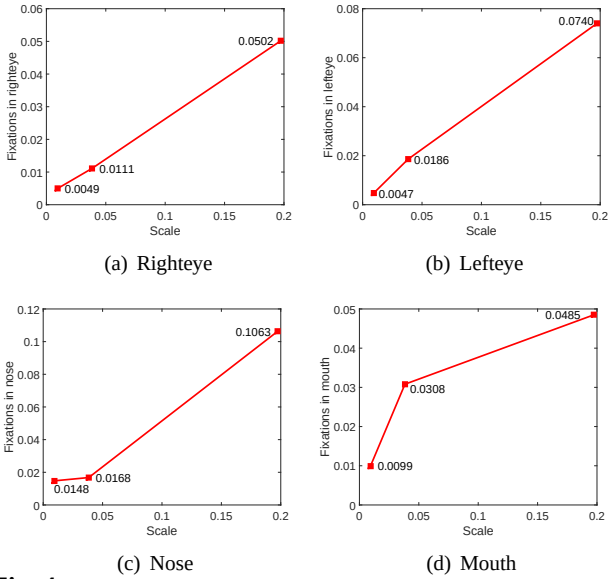
3. THE PROPOSED METHOD

This section mainly works on the proposed method for modeling saliency on face and facial features at different scales. In Section 3.1, we discuss preprocessing on the fixations for learning GMM. Next, GMM is learnt from the preprocessed training fixations, to be discussed in Section 3.2. In Section 3.3, we present the saliency detection method with the learnt GMM.

3.1. Preprocessing

For learning GMM, preprocessing has to be conducted to calibrate and normalize the eye fixations. Specifically, to avoid the uncertainty of face positions in different images, all fixations belonging to face region have to be calibrated in the following way.

As seen from Figure 5, Point A, the upper left point of PDM, is set to be the original point of the fixation coordi-



(a) Righteye (b) Lefteye

(c) Nose (d) Mouth

Fig. 4. Statistical results on proportions of eye fixations belonging to different facial features at various scales. Note that the results are averaged over all 281 images.

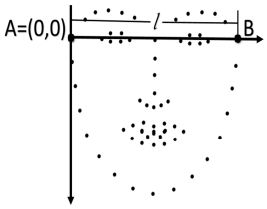


Fig. 5. Coordinate calibration and normalization on 68-point PDM.

nates in the face. Then, the coordinates (x, y) of fixations are calibrated to be (x^*, y^*) via translation:

$$\begin{cases} x^* = x - x_A \\ y^* = y - y_A, \end{cases} \tag{1}$$

where (x_A, y_A) is the coordinate for Point A.

Next, to deal with varying sizes of faces and facial features, fixations need to be normalized based on the width of face. To be more specific, the Euclidean distance l between Points A and B (as shown in Figure 5) is calculated as the unit length for fixation coordinates. As such, the normalized coordinates (x', y') can be calculated as follows,

$$\begin{cases} x' = \dfrac{x^*}{l} \\ y' = \dfrac{y^*}{l}. \end{cases} \tag{2}$$

Finally, the positions for eye fixations attended to faces can be represented in a uniformed coordinate system. This way, all fixations in faces from different images can be processed together for learning GMM.

3.2. Learning GMM

As aforementioned, the facial features attract a large amount of human attention, once the face is of large scale. Therefore, we can use the GMM to model the saliency within a face region, which has large saliency values around facial features.

Assuming that $\mathbf{x} = (x', y')$ is the calibrated and normalized coordinate of point (x, y) within a face, the GMM can be written as a linear superposition of Gaussian components in the form:

$$\sum_{k=1}^{K} \pi_k \mathcal{N}_k(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \tag{3}$$

and

$$\mathcal{N}_k(\mathbf{x}) = \exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{x} - \boldsymbol{\mu}_k)\}, \tag{4}$$

where π_k, $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$ are the mixing proportion, mean, and variance of the k-th Gaussian component. In (3), K is the total number of Gaussian components.

In fact, the GMM can be learnt from fixations of eye tracking data. Here, the EM algorithm [20] is applied to learn the GMM on the calibrated and normalized fixations falling into face regions. Note that varying numbers of Gaussian components are applied to different face scales. For example, as pointed out in Section 2, three facial features (right eye, left eye, and nose) tend to attract more attention when the face is in a large scale. Thus, the number of Gaussian components may be three, corresponding to the number of these facial features.

3.3. Saliency detection

Given the learnt GMM, the top-down conspicuity map on face channel ($\mathbf{F}$), denoted by $S(\mathbf{F})$, can be worked out on the basis of (3) and (4). However, for saliency detection the mean values $\boldsymbol{\mu_k}$ in (3) and (4) are replaced by the central points of facial features, when the number of Gaussian components is 3. This is because there may exist the deviation between the statistical centroids of Gaussian components and the detected central points of facial features. Note that the face detection method is mentioned in Section 2.2.

Next, similar to [15], the top-down conspicuity map is integrated with the bottom-up conspicuity maps of color ($\mathbf{C}$), intensity ($\mathbf{I}$), and orientation ($\mathbf{O}$). As a result, the final saliency map $\mathbf{M}$ can be generated by

$$\mathbf{M} = w_1 S(\mathbf{C}) + w_2 S(\mathbf{I}) + w_3 S(\mathbf{O}) + w_4 S(\mathbf{F}), \tag{5}$$

where $S(\cdot)$ is the normalized conspicuity map on each feature channel. $S(\mathbf{C})$, $S(\mathbf{I})$, and $S(\mathbf{O})$ can be obtained by the method in [9], whereas $S(\mathbf{F})$ needs to be yielded upon the learnt GMM as aforementioned. In addition, (w_1, w_2, w_3, w_4) are weights corresponding to each feature channel. They can be computed by least square fitting. For more details on computing these weights, refer to [19]. Figure 6 shows an example of overall procedure on our GMM-based saliency detection method.

4. EXPERIMENTAL RESULTS

In this section, experimental results are presented to evaluate the saliency prediction performance of our method. In Section 4.1, we provide the training results on learning GMMs from ground-truth fixations. In Section 4.2, we show the testing results of our method, in comparison with the conventional

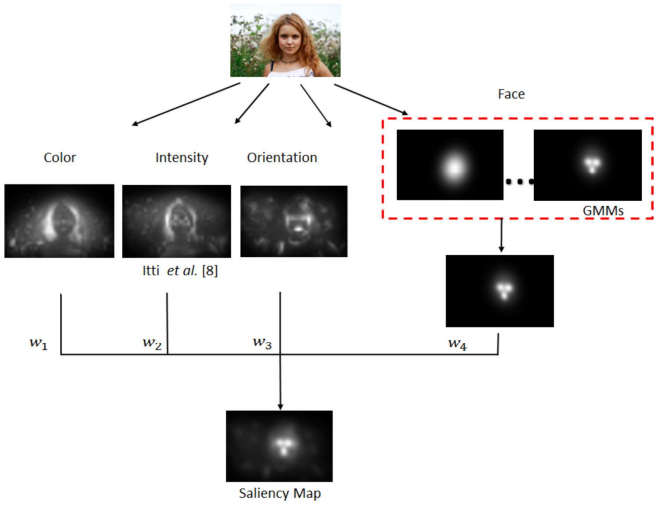


Fig. 6. Procedure of our GMM-based saliency detection method.

Table 1. The numbers of training faces and fixations

Training Results	Scale 1	Scale 2	Scale 3
Range	0-0.02	0.02-0.07	0.07-0.60
Training faces	403	116	27
Fixations for Training	8402	4441	3134

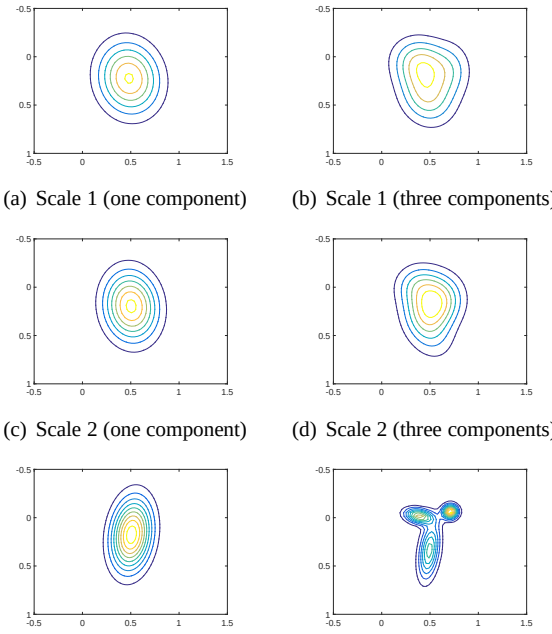
methods: Itti *et al.* [8], Cerf *et al.* [15], Zhao *et al.* [19], Judd *et al.* [13], Duan *et al.* [11], and Zhang *et al.* [12]. Here, the Area Under Curve (AUC), Kullback-Leibler (KL) [25] divergence, and the Normalized Scanpath Saliency (NSS) [26] are used as metrics for evaluating the accuracy of saliency detection.

4.1. Training Result

In our experiment, we divided our eye tracking database (presented in Section 2.1) into training and test sets. For training set, 401 images containing 546 faces were selected with 15,977 fixations. For the test set, the remaining 75 images with 75 faces were utilized, in which 3,268 fixations were obtained. Note that only 621 among 823 faces in 476 images can be correctly detected by the face detection method of Section 2.2. Thus, only 621 faces were used for training and testing for our experiments. Besides, there is no overlap between the training and test sets.

Since our method works on different scales of faces, we clustered 621 faces into three groups (546 faces for the training and 75 faces for test images), according to the corresponding face scales. Table 1 reports the ranges of these scales. Besides, Table 1 shows the numbers of faces and their corresponding fixations used for learning GMM at different scales. Note that there may be more than one scales of faces in these images. Thus, only the face numbers are provided in Table 1.

With the training eye tracking data, we learnt the GMMs for each face scale. The results of GMMs are shown in Figure 7. Note that the results for different numbers of Gaussian components are provided in this figure. As seen from Figure 7, for different numbers of components on GMMs, Scales 1 and 2 have the similar distributions of fixations. However, for Scale 3, the GMM with three components reflects that



(a) Scale 1 (one component) (b) Scale 1 (three components)

(c) Scale 2 (one component) (d) Scale 2 (three components)

(e) Scale 3 (one component) (f) Scale 3 (three components)

Fig. 7. Contours of GMMs learnt in our experiment.

fixations cluster in the facial features. Thus, to simplify our model, only one component is for Scale 1 and Scale 2, while three Gaussian components are applied for learning GMM at Scale 3. Moreover, the learnt GMMs of Figure 7 differ from Gaussian distributions of Cerf *et al.* [15] in the following three aspects:

- For Scale 3, three Gaussian components model the fixations around facial features.
- There exists the anisotropy in learnt GMMs, rather than the assumption on isotropy of Gaussian distribution [15].
- The standard deviation of the learnt GMMs decreases along with the increase of face scales, while the Gaussian distribution in [15] is simply assumed to be proportional to the face size.

4.2. Testing Result

AUC. In Table 2, we tabulate the AUC results of all methods at each scale, in order to show the accuracy of saliency prediction. Here, the average AUC values on all 75 test images are provided. As seen from this table, our method outperforms all other methods. Especially, there is 0.072 improvement of AUC over Zhao *et al.* [19] at Scale 3, in which the bottom-up features and corresponding weights are the same as our method. Such a significant improvement is possibly due to the three Gaussian components around the facial features in our learnt GMM. Moreover, we show in Figure 8 the ROC curves of saliency prediction for Scale 3 by all methods. Also, we can see that the ROC curve of our method is superior to the conventional ones.

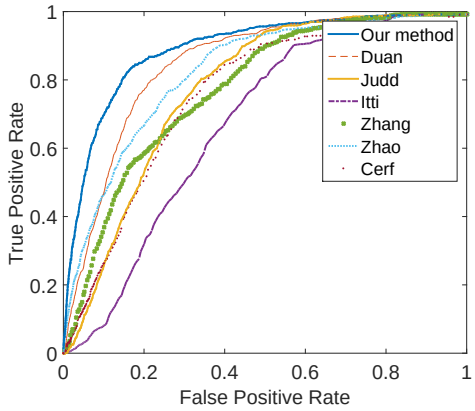


Fig. 8. Average ROC curves for all Scale 3 face images.

The KL divergence and NSS. In fact, the KL divergence measures the distance between distributions of detected and random saliency. NSS is computed to imply the correspondence between fixation locations and the saliency predictions. Methods with higher KL divergence and larger NSS value can better predict the fixations. The KL and NSS results of saliency detection by all methods are listed in Table 2. As can be seen from this table, our method again performs better than the conventional methods, especially for the large scale (i.e., Scale 3).

Saliency Map. Figure 9 shows the saliency maps detected by all methods. From this figure, we can see that our method can well locate the saliency regions. For small scales, it outputs the result similar to Zhao [19], but with more reasonable Gaussian distribution. For the large scale, it can yield the appropriate map, which well reflects the saliency regions of facial features.

5. CONCLUSION

In this paper, we have proposed to learn GMM from training fixations, for the top-down feature of face, on saliency detection of face images. Then, combined with conventional bottom-up features (i.e., color, intensity, and orientation), our saliency detection method was developed upon the learnt GMM. Different from other state-of-the-art methods on utilizing faces as the top-down feature for saliency detection, our GMM based saliency detection method benefits from the learnt GMM to precisely model the saliency values of face and even facial feature regions.

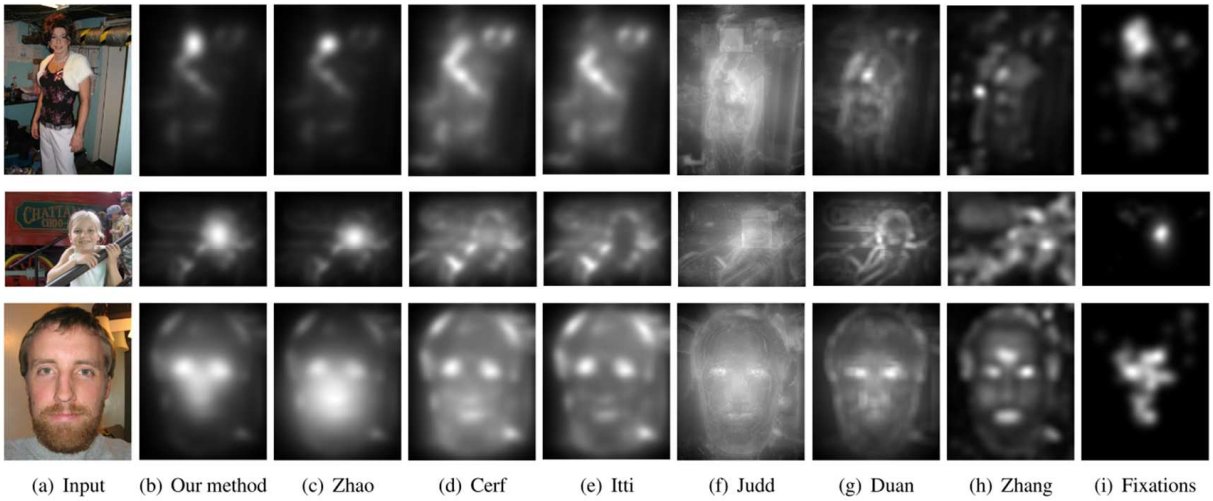
To facilitate saliency analysis of face and facial regions, we established a database of 476 images with 823 faces from the existing databases. Working on our database, GMMs were learnt for top-down feature in saliency detection. Finally, we evaluated our method by the AUC, KL divergence, and NSS metrics. In these three metrics, our method performed better than other conventional methods, especially when the face is with large scale.

References

[1] Ethel Matin, "Saccadic suppression: a review and an analysis.," *Psychological bulletin*, vol. 81, no. 12, pp. 899, 1974.

Table 2. The Results of AUC, KL and NSS

Methods	Scale 1			Scale 2			Scale 3			Average		
Metrics	AUC	KL	NSS	AUC	KL	NSS	AUC	KL	NSS	AUC	KL	NSS
Our Method	**0.773**	**0.220**	**1.578**	**0.812**	**0.300**	**1.953**	**0.868**	**0.372**	**1.897**	**0.797**	**0.263**	**1.720**
Itti[8]	0.748	0.206	0.966	0.743	0.219	0.863	0.719	0.207	0.842	0.743	0.209	0.922
Cerf[15]	0.763	0.215	1.101	0.785	0.245	1.179	0.776	0.233	1.075	0.770	0.225	1.117
Judd[13]	0.713	0.197	1.074	0.755	0.261	1.261	0.775	0.288	1.305	0.733	0.227	1.155
Zhao[19]	0.770	0.214	1.492	0.802	0.281	1.763	0.796	0.252	1.183	0.782	0.236	1.515
Duan[11]	0.746	0.193	1.025	0.804	0.260	1.275	0.843	0.304	1.510	0.775	0.226	1.160
Zhang[12]	0.723	0.168	0.895	0.752	0.209	0.960	0.759	0.226	1.037	0.736	0.187	0.932



(a) Input (b) Our method (c) Zhao (d) Cerf (e) Itti (f) Judd (g) Duan (h) Zhang (i) Fixations

Fig. 9. Saliency maps of different methods.

[2] Ali Borji and Laurent Itti, "State-of-the-art in visual attention modeling," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 1, pp. 185–207, 2013.

[3] Nicholas J Butko and Javier R Movellan, "Optimal scanning for faster object detection," in *CVPR*, 2009, pp. 2751–2758.

[4] Dashan Gao, Sunhyoung Han, and Nuno Vasconcelos, "Discriminant saliency, the detection of suspicious coincidences, and applications to visual recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 6, pp. 989–1005, 2009.

[5] Michael Rubinstein, Diego Gutierrez, Olga Sorkine, and Ariel Shamir, "A comparative study of image retargeting," in *ACM transactions on graphics (TOG)*. ACM, 2010, vol. 29, p. 160.

[6] Ulrich Engelke, Hagen Kaprykowsky, H Zepernick, and Patrick Ndjiki-Nya, "Visual attention in quality assessment," *Signal Processing Magazine, IEEE*, vol. 28, no. 6, pp. 50–59, 2011.

[7] Mai Xu, Xin Deng, Shengxi Li, and Zulin Wang, "Region-of-interest based conversational hevc coding with hierarchical perception model of face," *IEEE Journal of Selected Topics on Signal Processing*, vol. 8(3), 2014.

[8] Laurent Itti, Christof Koch, and Ernst Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[9] Dirk Walther and Christof Koch, "Modeling attention to salient proto-objects," *Neural networks*, vol. 19, no. 9, pp. 1395–1407, 2006.

[10] Jonathan Harel, Christof Koch, and Pietro Perona, "Graph-based visual saliency," in *NIPS*, 2006, pp. 545–552.

[11] Lijuan Duan, Chunpeng Wu, Jun Miao, Laiyun Qing, and Yu Fu, "Visual saliency detection by spatially weighted dissimilarity," in *CVPR*. IEEE, 2011, pp. 473–480.

[12] Jianming Zhang and Stan Sclaroff, "Saliency detection: a boolean map approach," in *ICCV*, 2013, pp. 153–160.

[13] Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba, "Learning to predict where humans look," in *ICCV*, 2009, pp. 2106–2113.

[14] Antonio Torralba, "Modeling global scene factors in attention," *JOSA A*, vol. 20, no. 7, pp. 1407–1418, 2003.

[15] Moran Cerf, Jonathan Harel, Wolfgang Einhäuser, and Christof Koch, "Predicting human gaze using low-level saliency combined with face detection," in *NIPS*, 2008, pp. 241–248.

[16] Guokang Zhu, Qi Wang, and Yuan Yuan, "Tag-saliency: Combining bottom-up and top-down information for saliency detection," *Computer Vision and Image Understanding*, vol. 118, pp. 40–49, 2014.

[17] Paul Viola and Michael Jones, "Rapid object detection using a boosted cascade of simple features," in *ICCV*, 2001, vol. 1, pp. I–511.

[18] Yan Hua, Zhicheng Zhao, Hu Tian, Xin Guo, and Anni Cai, "A probabilistic saliency model with memory-guided top-down cues for free-viewing," in *ICME*, 2013, pp. 1–6.

[19] Qi Zhao and Christof Koch, "Learning a saliency map using fixated locations in natural scenes," *Journal of vision*, vol. 11, no. 3, pp. 9, 2011.

[20] Todd K Moon, "The expectation-maximization algorithm," *Signal processing magazine, IEEE*, vol. 13, no. 6, pp. 47–60, 1996.

[21] Moran Cerf, E Paxon Frady, and Christof Koch, "Faces and text attract gaze independent of the task: Experimental data and computer model," *Journal of vision*, vol. 9, no. 12, pp. 10, 2009.

[22] Juan Xu, Ming Jiang, Shuo Wang, Mohan S Kankanhalli, and Qi Zhao, "Predicting human gaze beyond pixels," *Journal of vision*, vol. 14, no. 1, pp. 28, 2014.

[23] Subramanian Ramanathan, Harish Katti, Nicu Sebe, Mohan Kankanhalli, and Tat-Seng Chua, "An eye fixation database for saliency detection in images," in *Computer Vision–ECCV 2010*, pp. 30–43. Springer, 2010.

[24] Xiangxin Zhu and Deva Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *CVPR*, 2012, pp. 2879–2886.

[25] John A Hartigan and Manchek A Wong, "Algorithm as 136: A k-means clustering algorithm," *Applied statistics*, pp. 100–108, 1979.

[26] Robert J Peters, Asha Iyer, Laurent Itti, and Christof Koch, "Components of bottom-up gaze allocation in natural images," *Vision research*, vol. 45, no. 18, pp. 2397–2416, 2005.