Visual Attention Analysis and Prediction on Human Faces with Mole

Qianqian Wei, Guangtao Zhai, Chunjia Hu and Xiongkuo Min
Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai, China
{Mata-wqq, zhaiguangtao, hcj-sjtu, minxiongkuo}@sjtu.edu.cn

***Abstract*—Nowadays visual attention has been applied to many research and application problems. Different algorithms from low level to high level have been developed to detect the saliency map. For images with human face, high-level factors like mole may influence the visual attention. To investigate visual attention on human face with mole, we construct a Visual Attention database for Faces with Mole (VAFM) that contains face images, fixation density maps (FDM), landmark points as well as eye tracking data. Then we build visual attention model for face images with mole combining low-level saliency algorithms and high-level feature. Compared with the traditional low-level saliency algorithms, the proposed model perform better on our dataset.**

***Index Terms*—Saliency map, Human face, Visual attention, facial features, high-level factors**

I. Introduction

Visual attention is a special ability of human beings to observe the world. It can help people identify and distinguish all different characteristics in sight within seconds. It also draws our attention to some of the areas we see, which are called salient regions. In previous studies, visual attention has already been used to solve the signal processing problems widely (e.g., Automatic contrast enhancement [1]) and artificial intelligence application [2]. Visual saliency detection also has many practical applications in image processing and computer vision fields [3][4][5][6].

Many experimental results have already also been published based on the real testing experiments of visual attention. People classify visual attention simply as high-level and low-level. But in some specific cases involving high-level facts, there are few methods to deal with them effectively. For example, Birmingham et al. [7] point out that observers fixation are mainly derived by social information while the low-level saliency models do not account for those within social scenes and their important conclusion is that people tend to find out faces, especially at the eyes. Foulsham et al. [8] find out that hierarchy strongly predicted where participants focus on.

As described above, human face contains very important and classic high-level visual attention features, which provide abundant information. In the research of Vo ML [9], peoples attention is drawn to the eyes, nose or mouth when people gaze at others in the face. The fixation of those areas will be strengthened when the faces give signals from saliency area. For example, when people in the image start to talk and are opening their mouth, the subject will focus on the mouth region. However, the fixation on those areas will reduce when the prompt message is removed. Buchan et al. [10] found that observers are trying to achieve their goals by focusing more on particular parts of faces when they analyze the eye movement data. Subjects gaze more at eyes when judging emotions while they pay more attention to mouth to recognize words.

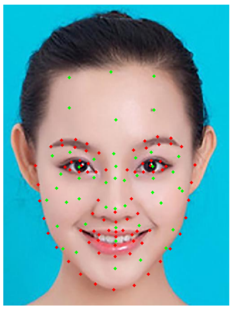

Fig. 1. Face landmarks and mole position on sample image.

Based on the works mentioned above, abundant high-level information from human faces has significant influence on visual attention. Researchers start to incorporate face cues into visual attention model. A large number of underlying algorithms are tested and certified valid, such as Judd [11], GBVS [12], IT [13] and etc. Most of the latest researches focus on the detailed characteristics of the faces, i.e., emotion [14], face size [15] and etc. In this paper, we pay our attention on the moles of human faces. As a high level factor on face, mole is one of the key characters in face recognition. Due to the salient color of the moles, they usually attracts much attention. Especially for images like selfies and identification photos, where faces occupy the scenes, moles become an important factor on the saliency map. Therefore, in this paper, we investigate the allocation of visual attention on faces with moles. At the beginning we build a Visual Attention database for Faces with Moles (VAFM). The database contains 258 images of faces with moles in different regions, as well as the corresponding eye-tracking data. Face detection and facial landmarks localization results are also available. Then we analyze the collected eye-tracking data. We define the area of interest (AOI) and calculate the salient distribution. Experimental results indicate that the models in different area have different impacts on people's attention. Thus we design a new visual attention model specifically for face image.

The rest of the paper is organized as followed: Section II introduce the eye-tracking experiments and some specifications of the VAFM database. Section III analyzes fixation distribution on faces. In section IV, high-level features are combined with low-level saliency computed by state-of-the-art saliency models and we propose the improved saliency model for faces with moles. The final section gives the concluding remarks.

II. Eye-tracking Experiments

A. Stimuli

We take 500 standard pictures. All test pictures are cropped to resolutions of 640*480 pixels, each of which contains a single human face out of 10 pairs of male and female sample faces. For every picture, the face region is divided into 43 different areas, and in certain areas we mark a point as mole with fixed gray value. Then we set another 6 original pictures of the same person without moles as a control group. With all 500 images (400 images with moles, 60 images without moles), we separate them into 5 sessions randomly.

B. Apparatus

In order to record the eye movement data, we use Tobii T120 Eye Tracker to perform subjective experiments. Tobii T120 has a 17 inch display with the resolution 1280*1024 pixels. It can record eye movements and perform preprocessed data at a sampling rate of 60Hz and 120Hz. Taking treatment efficiency into account, we adopt 60Hz sampling rate in our experiment, Tobii T120 has an effective tracking range of 50-80 cm as we expected.

C. Procedure and Test Condition

We recruit 5 pairs of male and female as experiment participants and each participant will observe the images in all 5 sessions, with Tobii T120 Eye Tracker tracking their eye movements. In each session, the test images show up randomly and stays on screen for 4 seconds, followed by a gray screenmask for 1 second between two images.

D. The VAFM Database

The Visual Attention database of Face with Mole is based on those data of eye-movement of testers. It also includes face images and fixation maps. Fixation density maps (FDMs) and facial landmarks are also available with this database. Facial landmarks is based on the program of FACE++ [16]. Fig. 1 is a sample image and shows the landmarks results with FACE++ [16] and relative positions of moles on face. The red points describe face region and green points denote moles' positions.

1) Fixation Map and FDM: Fixation maps and fixation density maps are generated from the eye-movement data, i.e., FixationDuration, MappedFixationPointX and MappedFixationPointY. We overly all fixation data into a binary map, in which the fixation positions are set to 1 while other positions are 0. And then they are filtered with a Gaussian kernel, generating FDM. Similar to [15], the standard deviation of the kernel is set to around 1 degree of visual angle.

2) Facial Landmark: Facial landmark is based on the program of FACE++ [16], one of the most famous human face recognition online application website. It can provide 83 landmark localization and points after delivering face image with computationally efficiency, which can be used to identify face region.

III. Visual Attention Analysis on Human Faces

A. Area of Interest Salient Distribution

We define the circle area around the moles, which occupies 1% of the whole image area as our Area of Interest (AOI). We calculate the time T_1 that people's attention stays on the area $f(AOI)$ in experiment groups and the time T_2 on the same coordinate in control groups. The salient value of corresponding area is defined as the difference between the two groups:

$$\begin{aligned} & f(AOI) = T_1/T_1' - T_2/T_2' \\ & 0 \le T_1, T_1', T_2, T_2' \le 1 \end{aligned} \tag{1}$$

where T_1' denotes the total time people spend on the image in experimental groups and T_2' denotes the time in control groups, to exclude the time people looks somewhere else other than photos.

B. Eye-movement Data Analysis

Based on the experimental results, we conclude as following:

- Salient value $f(AOI)$ varies on different mole areas, which indicates that the moles in different areas have different impacts on people's attention;
- The average salient distribute value on the right side is 0.0913, which is higher than 0.1178 on the left. That is to say, people pay more attention on the moles on the left side of face than the right side. Supposedly, stimulation on the left side of face draws more attention from people than that on the right side;
- As we consider the fact that female are paying careful attention to beauty than male. It is expectable that facial features such as mole are stronger attractors to female subjects than male. Besides, those facial features are more obvious and conspicuous on female's faces than on male's too. To analyze the overall fixation distribution with gender, we separated eye movement data from male and female subjects into two groups. The results lead to two conclusions: The moles on female faces draw more attention than moles on male faces; the sex of tester also has influences that female participants pay significantly more attention on moles than male participants.

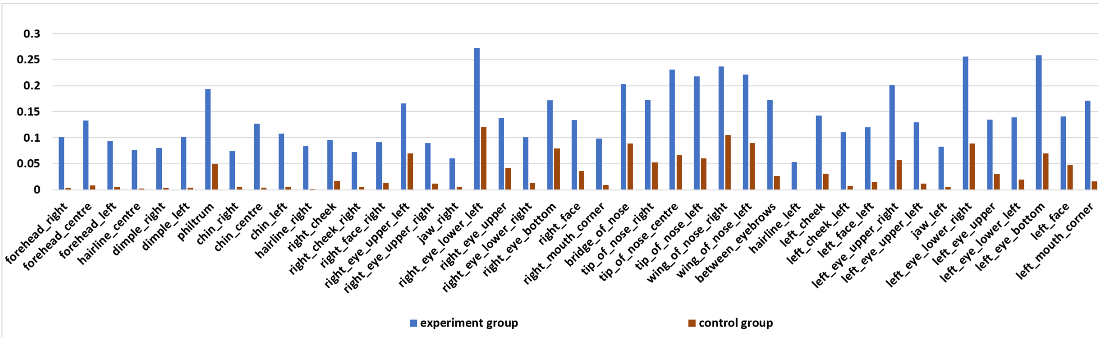

Fig. 2. Details about fixation distribution

More details about fixation distribution are illustrated in Fig. 2, where the blue cylinders denote the normalized average recorded time on AOI in experiment groups (T_1/T_1') and the orange ones show the normalized time in control groups (T_2/T_2').

IV. Visual Attention Prediction on Human Faces

As described in previous sections, traditional bottom-up visual saliency models preform not so well for face images with high-level factor like mole. So we design a new saliency model concentrating on faces with moles. Based on the visual attention analysis in Section III, we combine mole feature with low-level saliency. We will discuss some details of feature extraction and integration. The total efficiency of our improved algorithm are verified with VAFM database.

A. Feature Extraction

1) Low-level Saliency: Abundant of computational models are proposed to predict fixations. We select 4 typical state-of-the-art saliency algorithms to model low-level saliency, which are IT [13], AIM [17], GBVS [12] and SR [18].

2) High-level Facial Features: As described in previous section, mole is a strong attractor to subjects. We extract feature map for all those regions. First, we separate human face into 43 regions, which are agreed with mole position in Section III. The facial feature maps are calculated by simply placing uniform Gaussian kernels at the positions of mole. Feature map for mole, whose position is on region k, can be characterized by:

$$S_k = a_k * [M] \qquad 0 \leq k \leq 43 \qquad (2)$$

where a_k is corresponding weighting on region k; [M] is GAUSSIAN mask, which denotes the update region after Gaussian filter.

B. Feature Integration

In this section, we combine low-level saliency algorithm with extracted feature mole. The improved saliency method is characterized by:

$$S = W_1 * [Y] + W_2 * [aM] \qquad (3)$$

where [Y] is the saliency map calculated by low-level bottom-up algorithm, [aM] means GAUSSIAN mask with corresponding weighting.

C. Experiment Validation

1) Model Training: To determine the parameters W_1, W_2, we train a classifier from VAFM database using a 30-fold cross validation. We select 198 training images and 60 test images in each validation. From each training image, we randomly select 30 samples from top 30% and bottom 70% salient area, making a training set of 5940 positive and 5940 negative samples. Then feature values are concatenated into a feature vector, which is used to train a model through liblinear support vector machine [19]. Table I list the model parameters for different low-level saliency. The sample choosing and training method are similar to [11].

TABLE I
LEARNED MODEL PARAMENTERS

Low-level Saliency	W_1	W_2
AIM	0.3725	0.8624
IT	0.3704	0.9757
$GBVS$	0.6970	0.4838
SR	0.8274	0.9554

2) Evaluation Metrics: Similar to [20], 4 saliency evaluation metrics are adopted: area under the receiver operating characteristic (ROC) curve (AUC), linear correlation coefficient (CC), normalized scanpath saliency (NSS) and shuffled version of AUC (sAUC). The ROC curve can be plotted as the threshold varies; CC computer the linear correlation coefficient between saliency map and FDM; NSS means the mean value of the normalized saliency map at all fixation points; sAUC wipe off the influence of center bias.

3) Experiment result: Table II lists the experiment results. We compare the performance of traditional low-level saliency and the improved models. Scores demonstrate that visual attention models can predict fixation significantly better when incorporating the proposed features. Fig.3 illustrates sample images, FDMs, Gaussian mask, saliency maps calculated from low-level saliency model and improved models.

The facial feature mole is very important and more high-level factors should be taken into consideration during the study of visual attention prediction.

TABLE II
PERFORMANCE OF TRADITIONAL AND CORRESPONDING IMPROVED SALIENCY MODELS

Models	AUC		sAUC		CC		NSS	
	TRAD	IMPR	TRAD	IMPR	TRAD	IMPR	TRAD	IMPR
AIM [17]	0.6329	0.6819	0.5426	0.5844	0.2248	0.3692	0.5072	1.0938
IT [13]	0.5105	0.5889	0.5408	0.5856	-0.0895	0.1500	-0.1116	0.7193
SR [18]	0.6332	0.6788	0.5390	0.5827	0.2505	0.3725	0.6689	1.1564
GBVS [12]	0.6872	0.7262	0.5402	0.5818	0.3719	0.4837	0.8470	1,4102

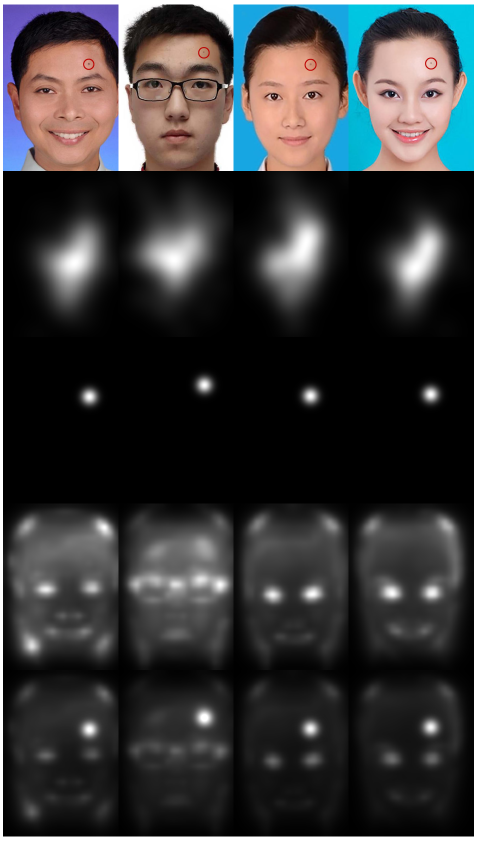

Fig. 3. Saliency maps of sample images: sample images, FDMs, Gaussian mask, saliency maps calculated from low-level saliency model GBVS[12], improved models with GBVS.

V. CONCLUSION

In this paper we study on the visual attention analysis and prediction on human face with high-level factor mole. We perform eye tracking experiments and build visual attention database named VAFM. With analyzed eye fixation distribution, we find out that mole in different area has different salient distribution. Through face detection and facial landmark localization, we separate human face into regions. Through learning, processed features are further integrated with low-level saliency models. Experimental results demonstrate that visual attention models perform much better on human fixations prediction when incorporating mole factor. The importance of high-level facial information has been proved.

ACKNOWLEDGEMENT

This work was supported in part by the National Science Foundation of China under Grants 61422112, 61371146, 61521062, 61527804, and National High-tech R&D Program of China under Grant 2015AA01590.

REFERENCES

[1] K. Gu, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, “Automatic contrast enhancement technology with saliency preservation,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 9, pp. 1480–1494, 2015.

[2] C. Siagian and L. Itti, “Biologically inspired mobile robot vision localization,” *Robotics, IEEE Transactions on*, vol. 25, no. 4, pp. 861–873, 2009.

[3] K. Gu, G. Zhai, W. Lin, and X. Yang, “Visual saliency detection with free energy theory,” *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1552–1555, 2015.

[4] X. Min, G. Zhai, Z. Gao, and K. Gu, “Visual attention data for image quality assessment databases,” *Proceedings - IEEE International Symposium on Circuits and Systems*, pp. 894–897, 2014.

[5] K. Gu, S. Wang, H. Yang, and W. Lin, “Saliency-guided quality assessment of screen content images,” *IEEE Transactions on Multimedia*, vol. 18, no. 6, pp. 1098–1110, 2016.

[6] X. Min, G. Zhai, C. Hu, and K. Gu, “Fixation prediction through multimodal analysis,” in *2015 Visual Communications and Image Processing (VCIP)*. IEEE, 2015, pp. 1–4.

[7] E. Birmingham, W. F. Bischof, and A. Kingstone, “Saliency does not account for fixations to eyes within social scenes,” *Vision research*, vol. 49, no. 24, pp. 2992–3000, 2009.

[8] T. Foulsham, J. T. Cheng, J. L. Tracy, J. Henrich, and A. Kingstone, “Gaze allocation in a dynamic situation: Effects of social status and speaking,” *Cognition*, vol. 117, no. 3, pp. 319–331, 2010.

[9] M. Vo, T. Smith, P. Mital, and J. Henderson, “Do the eyes really have it,” *Dynamic allocation of*, 2012.

[10] J. N. Buchan, M. Paré, and K. G. Munhall, “Spatial statistics of gaze fixations during dynamic face processing,” *Social Neuroscience*, vol. 2, no. 1, pp. 1–13, 2007.

[11] T. Judd, K. Ehinger, F. Durand, and A. Torralba, “Learning to predict where humans look,” in *Computer Vision, 2009 IEEE 12th international conference on*. IEEE, 2009, pp. 2106–2113.

[12] J. Harel, C. Koch, and P. Perona, “Graph-based visual saliency,” in *Advances in neural information processing systems*, 2006, pp. 545–552.

[13] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 11, pp. 1254–1259, 1998.

[14] H. Eisenbarth and G. W. Alpers, “Happy mouth and sad eyes: scanning emotional facial expressions.” *Emotion*, vol. 11, no. 4, p. 860, 2011.

[15] X. Min, G. Zhai, and K. Gu, “Visual attention on human face,” in *2015 Visual Communications and Image Processing (VCIP)*. IEEE, 2015, pp. 1–4.

[16] (2016) The FACE++ website. [Online]. Available: http://www.faceplusplus.com.cn/demo-landmark/

[17] N. Bruce and J. Tsotsos, “Saliency based on information maximization,” in *Advances in neural information processing systems*, 2005, pp. 155–162.

[18] X. Hou and L. Zhang, “Saliency detection: A spectral residual approach,” in *Computer Vision and Pattern Recognition, 2007. CVPR’07. IEEE Conference on*. IEEE, 2007, pp. 1–8.

[19] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, “Liblinear: A library for large linear classification,” *The Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.

[20] A. Borji, D. N. Sihite, and L. Itti, “Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study,” *Image Processing, IEEE Transactions on*, vol. 22, no. 1, pp. 55–69, 2013.