INDIVIDUAL TRAIT ORIENTED SCANPATH PREDICTION FOR VISUAL ATTENTION ANALYSIS

Aoqi Li and Zhenzhong Chen

School of Remote Sensing and Information Engineering,
Wuhan University, P.R.China

ABSTRACT

Scanpath refelects the shift of visual attention, therefore prediction of scanpath plays an important role in image analysis and understanding. However traditional scanpath prediction methods ignore the individuality of subjects such as oculomotor bias and other relevant factors. Hence, to make the scanpath prediction more accurate, we incorporate individual traits into a universal scanpath prediction framework for the subject based on the saccade distribution and factor weighting. Experiments demonstrate that our model improves the predicting performance, which proves that individuality is an important factor in scanpath prediction.

Index Terms— visual attention, oculomotor bias, individuality, scanpath prediction

1. INTRODUCTION

Visual attention can give us some intuition about the working mechanism of human visual system, thus enabling us to explore the cognitive strategies in visual search [1, 2] and object recognition [3, 4, 5]. In the past few decades, researchers have been committed to the modelling of image saliency maps [6, 7, 8, 9, 10], which can tell us which areas can draw our attention. However, saliency maps only show the spatial aspect of visual attention but discard its dynamic property. The shift of visual attention is reflected by scanpaths, which can manifest the order observers move their eyes over an image. To better understand the generation of human scanpaths, saccadic models have been proposed to tackle the problem of scanpath prediction. Wang *et al.* [11] first established a model to predict human scanpaths based on the information maximization criteria. Then a framework was proposed in [12] to integrate three main factors that affect human attention, i.e., low-level saliency, semantic content and spatial position. O. Le Meur *et al.* [13] shared a similar framework and estimated scanpaths by modelling bottom-up saliency, joint distribution of saccade amplitudes and orientations as well as memory effect and inhibition of return. In [14], some similar properties were found between super-Gaussian components (SGC) and fixations, and the process of generating a scanpath was transformed into selecting fixations with the maximum SGC responses. Reinforcement learning based method was also employed to learn scanpaths that resemble the ground truth data [15]. Obviously, all the aforementioned models assume that different observers have identical viewing behaviors when they are confronted with the same image.

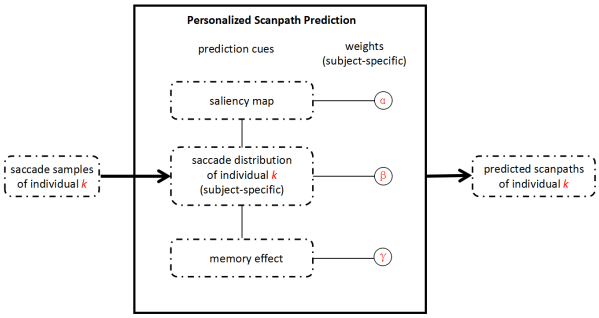


Fig. 1. The proposed framework for individual trait oriented scanpath prediction.

However, scanpaths vary greatly under different conditions. In [16], researchers found that individual scanpaths are idiosyncratic, i.e., more similar within an individual than between individuals. It has also been pointed out that human eye movements are not only spatially variant and scene category specific, but also differ from individual to individual [17] . Moreover, E. F. Risko *et al.* [18] have demonstrated that who a person is related to how they move their eyes. In addition to internal factors, distracting social stimuli can also affect how a person perceives an image and such effect is usually modulated by individual differences in sensitivity to social stimuli [19]. Therefore, we propose to take advantage of individual characteristics to make individual trait oriented scanpath prediction.

Inspired by [13], we consider scanpaths are generated by the interaction of three factors, i.e., image saliency, oculomotor bias and memory effect. The original universal framework is based on two assumptions: (1) different people have iden-

This work was supported in part by National Natural Science Foundation of China (No. 61471273), National Hightech R&D Program of China (863 Program, 2015AA015903), and Natural Science Foundation of Hubei Province of China (No. 2015CFA053).

tical oculomotor biases, so we only need to train a general distribution of saccade amplitudes and orientations for all the subjects; (2) the factors that influence scanpath generation are equally important. Nevertheless, such assumptions may not always stand to reason. In this paper, we introduce individual traits into a universal model based on the saccade distribution and factor weighting to predict scanpaths for a specific subject. The proposed framework is illustrated in **Fig. 1**. First individual saccadic distribution is explored and embedded in the prediction framework. Then the optimal weights of different prediction cues are determined for each subject to make individual traits oriented scanpath prediction.

The rest of the paper is organized as follows: Section 2 depicts individual saccadic traits. Section 3 gives a detailed illustration of the individual trait oriented scanpath prediction model. Experimental results and conclusions are presented in Sections 4 and 5, respectively.

2. INDIVIDUAL TRAITS IN SACCADE AMPLITUDES AND ORIENTATIONS

In the universal scanpath prediction framework, scanpaths are affected by the combination of image saliency, oculomotor bias and memory effect. The selection of the next fixation is based on the following formula:

$$x^* = \underset{x\in\Omega}{\operatorname{argmax}}\, p_S(x)p_D(d,\phi)p_M(x) \tag{1}$$

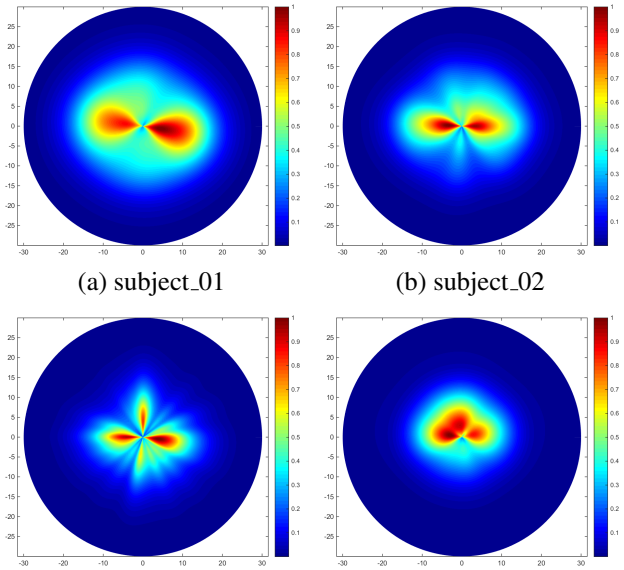
where x is a vector indicating the position of the predicted fixation, Ω is the set of points, $p_S(x)$ and $p_M(x)$ represent image saliency and memory effect respectively, while oculomotor bias $p_D(d,\phi)$ is modelled by the joint distribution of saccade amplitude d and orientation ϕ of all the observers.

Considering individual differences, we propose to adapt the universal framework to individual trait oriented scanpath prediction by substituting individual joint distribution of saccade amplitudes and orientations for the general distribution. The individual traits in saccades are described as follows:

$$p_D^k(d,\phi) = \frac{1}{n}\sum_{i=1}^{n} K(d-d_i^k, \phi-\phi_i^k) \tag{2}$$

where (d_i^k, ϕ_i^k) are saccade samples from individual k. $K(\cdot)$ is a two-dimensional Gaussian kernel, and the kernel bandwidth is tuned for each subject.

We choose four subjects from a public eye tracking data set (KTH Kootstra data set [20]) for illustration. Each subject viewed 99 images, of which 59 images are used for training, 20 images for tuning parameters and 20 images for testing. To compare the differences among individuals we normalize the probability obtained by equation (2) so that the maximal value for any individual is 1. **Fig. 2** shows the individual traits in saccade amplitudes and orientations. With regard to amplitudes, it is evident that subject_01, subject_02 have longer saccades than other subjects. As for orientations, we can find that:



(a) subject_01 (b) subject_02

(c) subject_03 (d) subject_04

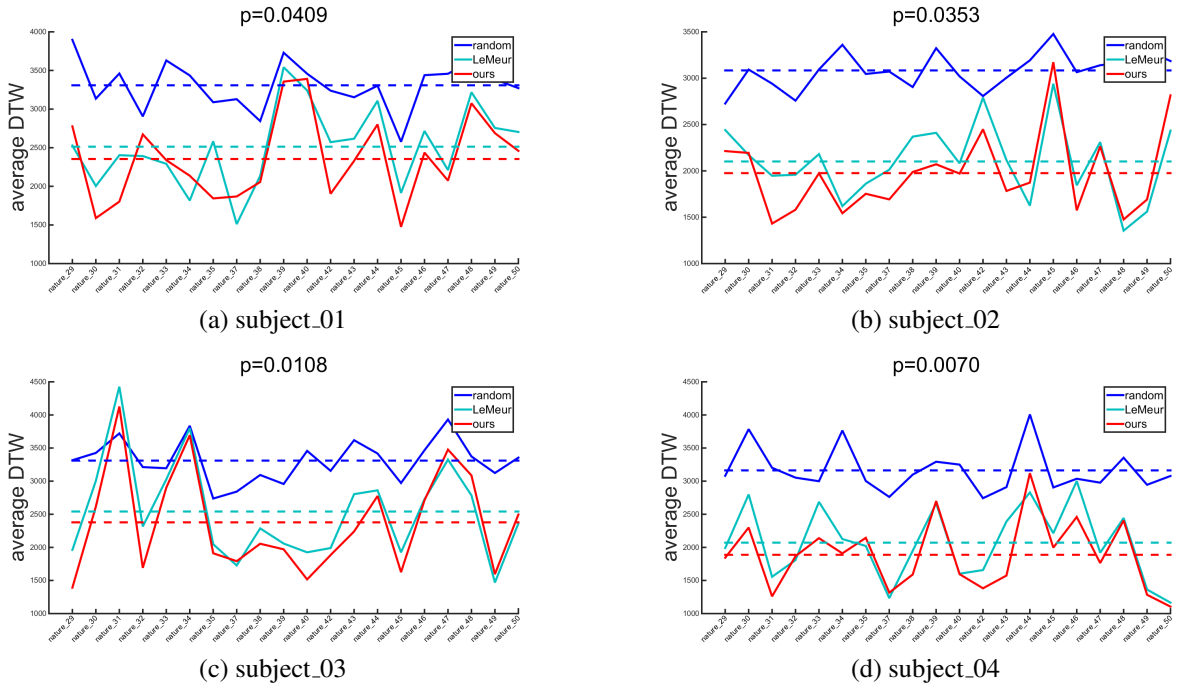
Fig. 2. Individual joint distribution of saccade amplitudes and orientations expressed in a polar coordinate system. Radial position indicates saccadic amplitude. Angle from the rightward horizontal axis indicates saccadic orientation.

- subject_01 and subject_02 prefer horizontal saccades.
- subject_03 tends to make both horizontal and vertical saccades.
- subject_04 is more likely to move their eyes upward than downward.

These differences in saccade amplitudes and orientations reflect oculomotor traits at the individual level. Hence, we intend to embed such individuality into the universal framework to make individual trait oriented scanpath prediction.

3. INDIVIDUAL TRAIT ORIENTED SCANPATH PREDICTION

As Equation (1) suggests, the universal scanpath prediction model does not differentiate the relative importance of the three factors that influence scanpath generation. Nevertheless, these cues to predict scanpaths should not just be treated equally. In [15], least square policy iteration was used to learn the weights of different features for scanpath prediction. Those features were summarized into four main categories, namely center bias, low-level image features, semantics and spatial distribution of eye fixation shifts. [15] divided visual exploration into different stages and found that weights of different cues are stage-specific, which means the weights vary



(a) subject_01 (b) subject_02

(c) subject_03 (d) subject_04

Fig. 3. Comparison among the random selection model, the universal scanpath prediction model [13] and our individual trait oriented scanpath prediction model. Paired t-test has been conducted on the results of the universal model and our personalized model.

for each stage. Similarly, in our individual trait oriented scanpath prediction framework, these factors for scanpath prediction are likely to be subject-specific. Hence, after obtaining the individual joint distribution of saccade amplitudes and orientations, we can modify Equation (1) as follows:

$$x^* = \underset{x\in\Omega}{\operatorname{argmax}}\, p_S(x)^\alpha p_D^k(d,\phi)^\beta p_M(x)^\gamma \tag{3}$$

where $p_S(x)$ and $p_M(x)$ represent image saliency and memory effect respectively, while $p_D^k(d,\phi)$ is the individual oculomotor bias of individual k. α, β and γ are model parameters indicating the relative importance of the three prediction cues.

Image saliency $p_S(x)$ is computed by the classic saliency model GBVS [7]. Memory effect $p_M(x)$ forces the influence of Inhibition of Return to deteriorate with time, which can be simplified as a linear model:

$$p_M(x) = 1 - \sum_{x_t\in\Omega_{visited}} \frac{t}{T} K(x_t) \quad (1 \leqslant t \leqslant T) \tag{4}$$

where $\Omega_{visited}$ is the set of visited points, t is the time interval between the visited point and the fixation to be predicted, T is a parameter that determines how long a visited point will affect the following fixations, and $K(\cdot)$ is a two-dimensional Gaussian kernel.

Since human eye movement is a stochastic process, even two scanpaths of the same subject viewing the same image can not be identical. Therefore, when generating a scanpath, we do not choose the point with the maximal probability as Equation (3) suggests. Instead, we generate several random points as candidates, of which the one with the largest probability is selected as a fixation.

4. EXPERIMENT AND RESULTS

4.1. Experiment Setup

The experiment is conducted on the public KTH Kootstra data set [20]. This data set includes eye tracking data from 31 subjects, from which we choose 4 (subject_01, subject_02, subject_03, subject_04) for illustration. Each subject viewed 99 images displayed full-screen with a resolution of 1024 by 768 pixels on an 18 inch CRT monitor of 36 by 27 cm at a distance of 70 cm from the participants. So for each subject, we have saccade samples from 99 scanpaths, of which 59 are used for training the individual joint distribution of saccade amplitudes and orientations, 20 are used for tuning all the parameters in the proposed model, and the rest are used for testing. Since the framework takes into account the randomness of human eye movements, given an image, we generate 10 scanpaths for each subject. The average DTW score of the 10 scanpaths indicates the model performance on an image.

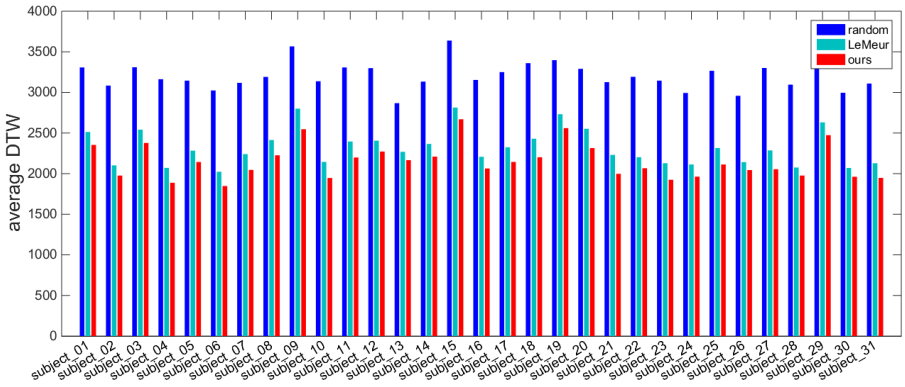


Fig. 4. Overall performance for all the 31 subjects evaluated by DTW scores. Three models are compared: the random selection model, the universal scanpath prediction model and our individual trait oriented scanpath prediction model. Paired t-test has been conducted on the results of the universal model and our personalized model ($p < 0.05$).

4.2. Results and Analyses

Dynamic Time Warping (DTW) [21, 22] is commonly used for measuring sequence similarity. For two scanpaths, we can calculate the DTW between them based on the following formula:

$$D(i,j) = \delta(i,j) + min \begin{Bmatrix} D(i-1,j-1) \\ D(i-1,j) \\ D(i,j-1) \end{Bmatrix} \quad (5)$$
$$(1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n)$$

where $\delta(i,j)$ is the Euclidean distance between the ith fixation of the first scanpath and the jth fixation of the second scanpath. D is the cumulative matrix. m and n are the length of two scanpaths respectively. $D(m,n)$ is the DTW distance between the two scanpaths.

In this part, we use the average DTW as the performance metric, a smaller value of which indicates a better performance. Given an image, we generate 10 scanpaths for each subject using three different models, i.e., the random selection model, the universal scanpath prediction model [13] and our individual trait oriented scanpath prediction model, then we compute the average DTW between the generated scanpaths and the actual scanpath. Statistical analysis has also been conducted. **Fig. 3** depicts different model performances on the 20 test images for four subjects and gives the corresponding p values. Both the universal scanpath prediction model [13] and our personalized model perform better than the random selection model. In **Fig. 3**, dashed lines represent the overall performance on the 20 test images, so averagely speaking, for the four subjects, scanpaths obtained by our personalized prediction model are closer to ground truth data, which means individual traits do play a key role in scanpath prediction. To further demonstrate the superiority of the proposed model, we conduct the paired t-test between the results of the universal scanpath prediction model [13] and our individual trait oriented model at the 5% significance level with the null hypothesis that the results of both methods are not significantly different (i.e., alternative hypothesis: the results are significantly different). Note that in Section 2 we have found subject_01 and subject_02 prefer horizontal saccades, which resembles the universal saccade tendency [13]. Therefore, the fact that the p values of the two subjects ($p_1 = 0.0409, p_2 = 0.0353$) are relatively larger than other two subjects ($p_3 = 0.0108, p_4 = 0.0070$) consolidates the position of individual traits in scanpath prediction.

Fig. 4 gives comparisons of the three models for all the 31 subjects. For each subject, the DTW score indicates the average performance on the 20 test images and paired t-test has been conducted ($p < 0.05$). We can conclude that our individual trait oriented scanpath prediction model achieves a better performance.

5. CONCLUSIONS

Most current researches on scanpath prediction proposed universal models, ignoring the individuality of different subjects. In this paper, we consider the differences at individual level to make individual trait oriented scanpath prediction. Our model introduces individual traits into a universal scanpath prediction framework, which describes scanpaths as the result of three factors, i.e., image saliency, oculomotor bias and memory effect. Individual traits are modelled from two aspects: (1) individual saccadic characteristics are modelled in the form of saccade amplitudes and orientations, and this personalized saccade distribution is incorporated into the original framework; (2) factor weights are introduced to adjust the relative importance of the prediction cues that are generally assumed equally important in scanpath prediction. Our model reflects the subject-specific property of scanpaths. Experiments demonstrate that our model achieves a better performance than the universal model [13], proving that individuality is an important factor in scanpath prediction.

6. REFERENCES

[1] R. Li, P. Shi, J. Pelz, C. O. Alm, and A. R. Haake, "Modeling eye movement patterns to characterize perceptual skill in image-based diagnostic reasoning processes," *Computer Vision and Image Understanding*, vol. 151, pp. 138 – 152, 2016.

[2] D. Hoppe and C. A. Rothkopf, "Learning rational temporal eye movement strategies," *Proceedings of the National Academy of Sciences*, vol. 113, no. 29, pp. 8332–8337, 2016.

[3] G. Ge, K. Yun, D. Samaras, and G. J. Zelinsky, "Action classification in still images using human eye movements," in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 16–23.

[4] Iván González-Díaz, Vincent Buso, and Jenny Benois-Pineau, "Perceptual modeling in the problem of active object recognition in visual scenes," *Pattern Recognition*, vol. 56, pp. 129 – 141, 2016.

[5] W. Zou, Z. Liu, K. Kpalma, J. Ronsin, Y. Zhao, and N. Komodakis, "Unsupervised joint salient region detection and object segmentation," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3858–3873, Nov 2015.

[6] L. Itti, C. Koch, and E. Niebr, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 20, pp. 1254–1259, 1998.

[7] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in Neural Information Processing Systems*, 2007, pp. 545 – 552.

[8] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 194–201, Jan 2012.

[9] A. Tal, L. Zelnik-Manor, and S. Goferman, "Context-aware saliency detection," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 34, pp. 1915–1926, 2012.

[10] A. Hornung, Y. Pritch, P. Krahenbuhl, and F. Perazzi, "Saliency filters: Contrast based filtering for salient region detection," *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 00, pp. 733–740, 2012.

[11] W. Wang, C. Chen, Y. Wang, T. Jiang, F. Fang, and Y. Yao, "Simulating human saccadic scanpaths on natural images," in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 441–448.

[12] H. Liu, D. Xu, Q. Huang, W. Li, M. Xu, and S. Lin, "Semantically-based Human Scanpath Estimation with HMMs," in *2013 IEEE International Conference on Computer Vision*, 2013, pp. 3232–3239.

[13] O. Le Meur and Z. Liu, "Saccadic model of eye movements for free-viewing condition," *Vision Research*, vol. 116, Part B, pp. 152 – 164, 2015, Computational Models of Visual Attention.

[14] X. Sun, H. Yao, and R. Ji, "What Are We Looking For: Towards Statistical Modeling of Saccadic Eye Movements and Visual Saliency," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1552–1559.

[15] M. Jiang, X. Boix, G. Roig, J. Xu, L. Van Gool, and Q. Zhao, "Learning to predict sequences of human visual fixations," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1241–1252, June 2016.

[16] N. C. Anderson, F. Anderson, A. Kingstone, and W. F. Bischof, "A comparison of scanpath comparison methods," *Behavior Research Methods*, vol. 47, no. 4, pp. 1377–1392, 2015.

[17] O. Le Meur and A. Coutrot, "Introducing context-dependent and spatially-variant viewing biases in saccadic models," *Vision Research*, vol. 121, pp. 72–84, 2016.

[18] E. F. Risko, N. C. Anderson, S. Lanthier, and A. Kingstone, "Curious eyes: Individual differences in personality predict eye movement behavior in scene-viewing," *Cognition*, vol. 122, no. 1, pp. 86 – 90, 2012.

[19] B. R. Doherty, E. Z. Patai, M. Duta, A. C. Nobre, and G. Scerif, "The functional consequences of social distraction: Attention and memory for complex scenes," *Cognition*, vol. 158, pp. 215 – 223, 2017.

[20] G. Kootstra, B. de Boer, and L. R. B. Schomaker, "Predicting eye fixations on complex visual stimuli using local symmetry," *Cognitive Computation*, vol. 3, no. 1, pp. 223–240, 2011.

[21] H. Sakoe and S. Chiba, "A dynamic programming approach to continuous speech recognition," *Proceedings of the Seventh International Congress on Acoustics*, vol. 3, pp. 65 – 69, 1971.

[22] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43 – 49, Feb 1978.