DOI: 10.2352/ISSN.2470-1173.2016.2.VIPC-230

Towards Region-of-Attention Analysis in Eye Tracking Protocols

Yingbin Wang, Xiu Chen and Zhenzhong Chen; School of Remote Sensing and Information Engineering, Wuhan University; Wuhan, China

Abstract

Eye tracking plays an important role in understanding human visual perception, especially visual attention and quality assessment. In this paper, we propose an efficient algorithm for identifying the region of attention in the eye tracking protocol. We use velocity-threshold based fixation identification algorithm to identify fixations and divide them into different clusters. In a particular cluster, we propose a densest position based method to locate the center of human attention. We have implemented the eye tracking protocols and conducted the experiments to evaluate the performance of our system. As shown in our results, the proposed method shows superior performance when compared to the traditional method.

Introduction

Eye tracking has been applied in many areas, such as neurology, psychology, marketing and advertising [1-4]. Eye movements are now recorded to analyze the behavior pattern of the individual. For example, it can help to study the behavior of the driver in driver training, and evaluate the effectiveness of advertisements [5-7]. Eye movements provide a channel to the recognition of individuals, which again stimulates the rise of commercial application in the relevant industries [8]. Compared with traditional approaches, the eye tracker provides more accurate data for the analysis of our mental mechanism and behavior habitats.

The eye movements contain a wealth of information about the observer's psychological activities. Many researchers have extensively explored the relations between eye movements and mental mechanism. We use a non-contact eye-tracker which measure the user's sight based on the pupil center cornea reflection technique[17]. It provides accurate raw gaze data about the individual's eye movements which is very important for further analyses.

The mainstream researches in eye tracking focus on how to acquire accurate data, filter or identify effective data for higher level analyses, and interpret eye movements as different individual's preference or mental activities. For spatial characteristics, the algorithm to identify fixations can be classified into three classes: velocity-based [9], dispersion-based [10], and area-based [11]. Typical algorithms identify fixations by applying velocity-threshold or limiting the size of the region, and partition the fixations into clusters according to the continuity of time [12-13]. Therefore, we combine the velocity based and dispersion based algorithms to identify fixations and generate accurate clusters of region-of-attention (ROA). Velocity-threshold fixation identification (I-VT) is a velocity-based method that is efficient to generate the cluster of fixations and has been adopted in practical applications. Dispersion based algorithms can identify ROA more efficiently. In a specified cluster of fixations, the mean position of the cluster is typically used to represent the center of the viewer's visual attention. However, the existing mean based method is sensitive to noise and vulnerable to the change of its parameters. Therefore, we proposed a method based on maximum density to locate the center of the cluster.

In this paper, we first utilize the velocity threshold fixation identification (I-VT) to identify fixations from the raw eye movements, and divide fixations into different clusters which direct towards ROAs. In a specified cluster, we discover the center of the viewer's visual attention according to the density of fixation. To obtain the raw eye movements data, we import a commercial eye tracker, and develop a system based on its SDK. Our developed system includes an integral process towards the raw eye movements we possess. We identify the fixations within them and partition the fixations into different ROAs. Finally, the centers of ROAs are connected to display the path of the viewer's browsing.

The paper is organized as follows. Section 2 explains the identification algorithms of viewer's visual attention. Section 3 performs the experiments to validate our algorithm. Section 4 draws the conclusions.

Visual Attention Identification

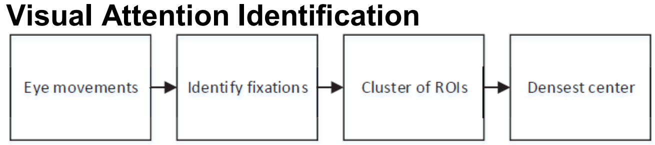


***Figure 1**. The framework of the system.*

The framework of visual attention identification in this paper is shown in Fig. 1. Given the raw eye movements collected from the eye tracker, we manage to dig up the information of the observer's attention during viewing stimuli. First, we perform the identification of fixations from saccades, which is regarded as the fundamental task and has long drawn the researchers' attention. Then, the fixations are partitioned into different clusters of ROAs. When looking through the stimuli for image information or interested things, the person's sight is likely to centralize around a particular ROA, and then move towards a new region [14]. Finally, we proposed a densest based method for the center of the cluster. The connection of these centers describes the path of the viewer's visual attention.

To implement the integral procedure of eye tracking and analyzing, we import Tobii X120 Eye Tracker, and develop a comprehensive system in C# based on its SDK. Our system consists of 3 modules, including design module, replay module, statistic module. It provides a convenient way for recording and analyzing the eye gaze data. Using the system, studies can be carried out in a timely and cost efficient way without the need for extensive training. The main algorithms in this system are explained as follows.

Identify Fixations

The recorded eye tracking data contains fixations, saccades, smooth pursuit, and blink. Fixations are the short stops over informative region of attention, and saccades are the quick movement between the fixations. Smooth pursuit is the kind of eye movement that follow a moving object, and it can be ignored when the stimuli are static. Blink is a rapid closing of the eyelid, which interrupts the continuity of eye tracking data.

Among these types of eye movements, the fixations and saccades are the most commonly used in the investigations of cognitive processes in human beings. And fixations contain more information about the viewer's attention than saccades. Primarily, we focus on the two major types of identification algorithms here: dispersion based algorithms, velocity based algorithms. Dispersion based algorithms utilize the fact that fixation groups have the tendency of gathering together, thus the dispersion is lower than a particular threshold. This method can be adopted to identify the regions of interest. Velocity based algorithms calculate the point-to-point velocities, then separate fixations and saccades by using the principle that fixations have lower velocity than saccades.

In this paper, we parse the eye tracking data into fixations and saccades by utilizing velocity based algorithm. Each eye tracking data contains the location and a timestamp, which can be represented as an triple, $< x_i, y_i, t_i >, \mathrm{i} = 1,2,3, \dots, \mathrm{n}$. It begins by computing the velocity of each point, in the protocol where the velocity corresponds to the distance between the current point and its successor in unit time. We can easily calculate the distance S_{i+1} and the interval $\Delta\mathrm{t}_{i+1}$ between the current point and the next one according to equation 1-3, where x_i, y_i, t_i refer to eye tracking data of point_i.

$$\mathrm{S}_i = \sqrt[2]{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}\,, \tag{1}$$

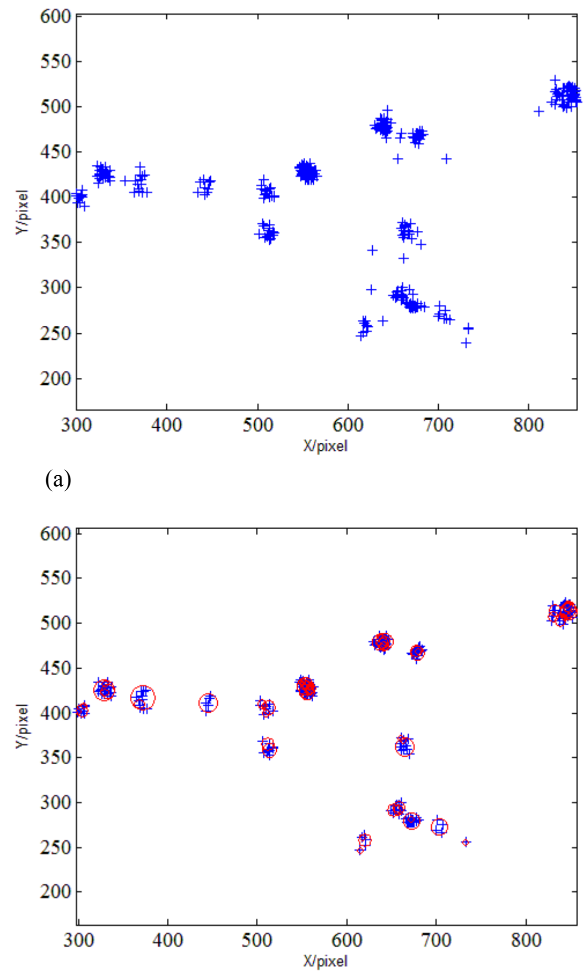
$$\Delta\mathrm{t}_i = \mathrm{t}_{i+1} - t_i, \tag{2}$$

$$\mathrm{V}_i = \mathrm{S}_i / \Delta\mathrm{t}_i, \tag{3}$$

We set a velocity threshold T_v to identify fixations from the saccades: the point belong to the same fixations are supposed to have lower velocities than the velocity threshold, while the saccades with higher speeds.

$$\begin{cases} \mathrm{P}_i \ \ belongs\ to\ fixation, \mathrm{V}_i < \mathrm{T}_v \\ \mathrm{P}_i \ \ belongs\ to\ sccade \quad, \mathrm{V}_i > \mathrm{T}_v \end{cases}, \tag{4}$$

The velocity based algorithm requires one parameter, the velocity threshold, which is related to the distance from eye to display screen and the characteristics of the stimuli on which we performed the experiments. Because of the high sampling frequency of the eye tracker we have, we tend to utilize a lower velocity threshold to perform a rough matching on the raw data. Then, the points with higher velocities than the threshold are regarded as the saccades and are discarded, meanwhile, the successive points are collapsed into several fixation clusters. The fixation clusters are marked with a new chronological order after discarding the saccades. In this process, we finally receive the initial fixation clusters in chronological order. The points in Figure 2. (a) shows the raw eye tracking data, most of the point tend to gather to groups, with a few point distributed among them. The result of fixation identification using I-VT algorithm are shown in (b). The points distributed among each group are regarded as saccades and discarded. And the initial points are collapsed into several fixation clusters. However, as the low threshold we used, some points within the same ROA are divided into two or more clusters.



(a)

(b)

***Figure 2**. The blue crosshairs represent the eye tracking data, and the crosshairs in the same circle are belonged to the same fixation groups. (a) shows the raw eye tracking data. (b) shows the result after performing I-VT, saccades are discarded and the points are separated into different fixation groups.*

Cluster of ROAs

Many researches in psychology require an effective recognition of areas of the observer's interests, and then specify the location, size or other attributes for the area [8, 15-16]. Therefore, clustering algorithms are needed to group fixations into clusters according spatial and temporal criterions.

Fixations gather within a particular ROA, but the saccades direct towards nothing. The continuity of fixations is divided by saccades. Therefore, the consecutive fixations are grouped into the same cluster after removing the saccades. The former process has produced the basic fixation clusters. Nevertheless, the velocity based algorithm adopted has some drawbacks for its lack of robustness. For example, when the eye moves slowly and the velocity threshold is not large enough, the above algorithm is very likely to produce a large cluster which group massive points together. To a certain extent, our low threshold avoids this circumstance, but it is probably to divide the fixations within the

DOI: 10.2352/ISSN.2470-1173.2016.2.VIPC-230

same ROA into two or more clusters. Therefore, we need to perform the further refinement.

As the Table 1 shows, we perform the dispersion based algorithm on the centroids of initial clusters coming from the previous step. We measure the distance between the current centroid and the next one, and compare it with the predefined dispersion threshold. Once the distance is smaller than what we expected, we continue to compare the current centroid with the following one. The consecutive clusters which meet the aforesaid requirements will collapse into one cluster. So far, the cluster we obtained each directs towards a particular region of attention (ROA). We suppose that the time of human paying attention to an object required at least 0.1 s. According to the sampling frequency of the eye tracker, we can calculate a number threshold to limit the minimum size of a cluster. Figure 3 shows the result of clustering. Compared to Figure 2. (b), the tiny cluster collapse into a larger cluster that better represent the region of attention.

Table 1: Pseudocode for clustering algorithm

Clustering algorithm (initial clusters, dispersion threshold, number threshold)
Descriptions: N = the number of initial clusters, $i \in [1, N-1]$.
Calculate the centroid of each cluster we received, the centroids are sorted in chronological sequence
For the centroid C_i
    $j = i + 1$
While $j < N$
Calculate the distance between the centroid C_i and C_j
If the distance is smaller than the dispersion threshold
    $j = j + 1$
Else
    $k = j - 1$
Merge the initial cluster from i to k into one cluster of ROA
    $i = j$
Endif
Endwhile
End
Check the number of the points in each cluster.
For the cluster in derived from the former step
If the number of the points in the cluster is smaller than the number threshold
    Discard the cluster
End
End
Return clusters of ROAs

Densest Center

After the process of the algorithms above, saccades are removed from the eye tracking data, and accurately clustered into several regions of interest. These algorithms above are efficient in finding regions of attention, but in the issue of appointing a center for a region, most of these approaches failed to achieve accurate results. Traditionally, the mean location of the fixations is calculated as the center, which means that each fixation in one cluster is treated equally without considering the difference among fixations.

In this paper, different algorithms are utilized to complete the identification and clustering of fixations, and on the basis of these regions-of-attention. We propose the measurement of center in considering of their densities. We all have the common sense that humans have a tendency to focus on their interested scenes, consequently, the eye movements are expected to be focused on that target region on the stimuli. Thus, the densities of fixations must to be of vital importance in the researches of human beings' visual attention.

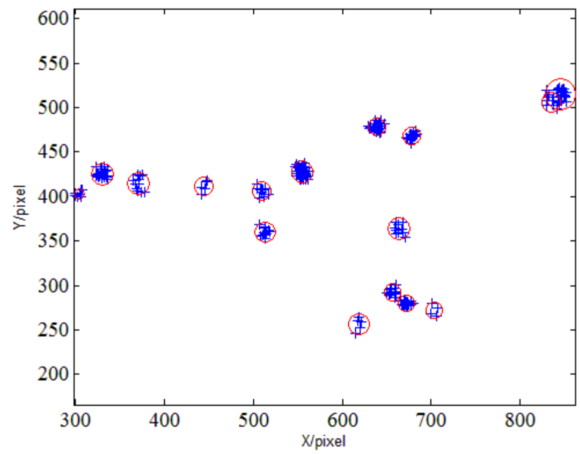


***Figure 3**. After clustering of ROAs, the initial fixation, which are divided into different groups, are clustered into the same ROAs.*

We proposed a novel method for the identification of visual attention. In terms of a particular cluster of ROA, we calculate the densities for each fixation therein and discover the maximum density value which point to the densest position. The densest fixation position is where the viewer put on the most attention in that region of attention.

In practical, the above algorithm can be expressed in the following way. Given a cluster which contains n fixations $\{f_1, \ldots, f_n\}$, we try to discover a location to represent the center of the region of attention. The eye movements in one cluster are characterized as $\{x_i, y_i, t_i, d_i\}$ where x_i and y_i denote the coordinates of the fixations f_i whose duration time is d_i, and the recording time is t_i. First, we need to compute a statistical vector whose elements are defined as:

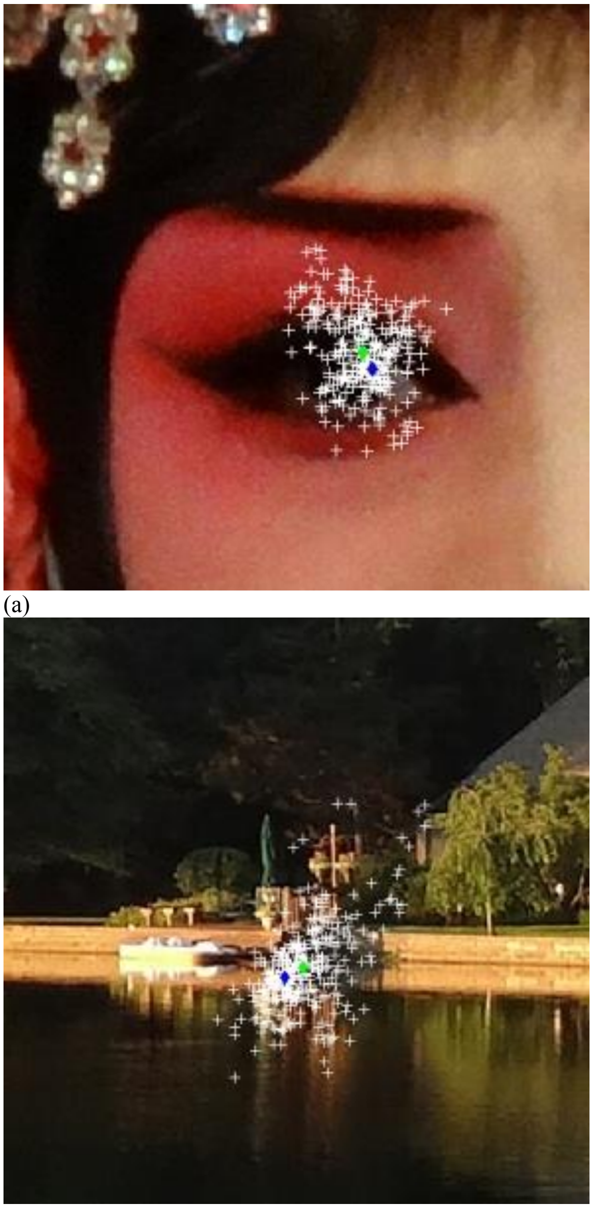
$$\text{density}(i) = \left\{\textstyle\sum_{j=1}^{n} d_i \,\middle|\, distance_{f_i,f_j} < threshold\right\}, \qquad (1)$$

Where $distance_{f_i,f_j}$ denotes the Euclidean distance between the fixation f_i and f_j, the fixation duration is added up once the distance is less than a predefined threshold. The fixation corresponding to maximum density is then chosen as the center:

$$\text{Center}(x, y) = \{(x_m, y_m) | density(m) = \max(density)\}. \qquad (2)$$

Figure 4 shows the result of locating the center by mean location and the densest position. The method we proposed shows better performance in locating the center we are interested at. As the mean location are easily affected by noise, it always failed to locate the center accurately.

DOI: 10.2352/ISSN.2470-1173.2016.2.VIPC-230



(a)

(b)

Figure 4. Utilizing Identified fixations (white) within the same region-of-attention, we compute the mean location (green) and the densest position (blue).

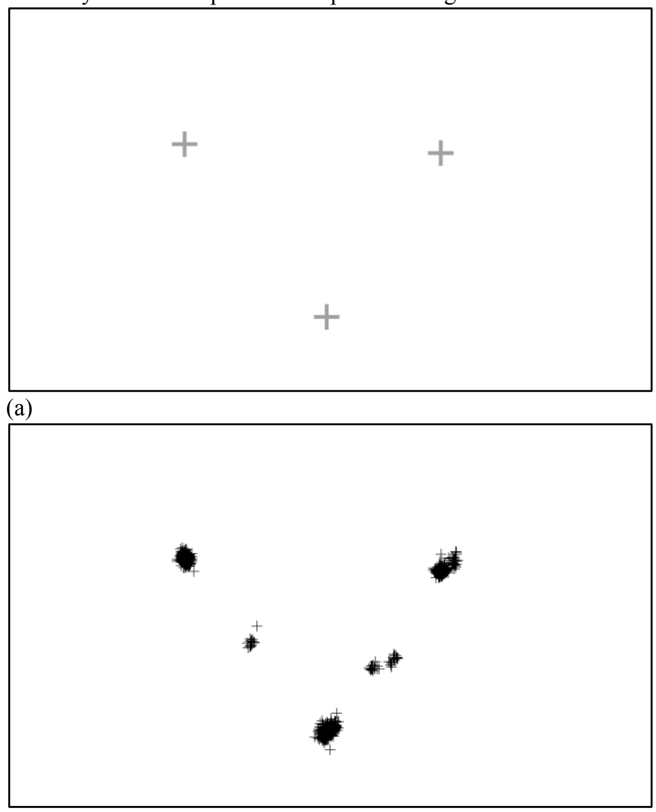
Experiments

Experiment Setup

The experiment is performed based on Tobii X120 Eye Tracker. Using the eye tracker, we can perform eye tracking tests in a convenient and efficient way. The eye tracking data includes the position, the timestamp and other related parameters. The eye tracker is placed between the monitor screen and the participant. In the experiments, we perform a calibration to ensure the precision of the eye tracker. The sampling rate of our experiments is set to 120 Hz, which will generate enough accurate data for later analyses.

Experiments and Analyses

The ground truth is uncertain in the experiments. Therefore, we predefine a pattern and instruct the viewers to fixate their attention on the center of the pattern. We perform our experiments on Figure 5. (a), testers are required to focus on the center of the crosshair.one by one. Figure. 5. (b) shows initial eye tracking data, each tiny crosshair represents the point tester gazed at.



(a)

(b)

Figure 5. (a) is the image for experiment. (b) shows the initial eye tracking data.

Experiments on fixation identification

We use different velocity thresholds to validate the effect of fixation identification. We evaluate the effect according to the number of the cluster, the percentage of remained points after the process, and the maximum variance of all cluster. The result is shown in Table 2.

As the velocity threshold reduce, more points are regarded as saccades and discarded, and more clusters are generated. But an extremely low threshold will result in large abandon of the initial points and reduction of clusters. The maximum standard deviation become smaller as the threshold decreases, and it keep stable when the threshold is high enough. In this paper, we use a low threshold to perform fixation identification, which helps discard the saccades from the initial data efficiently. In Figure 6, the initial eye tracking data are separated into different groups. However, various groups of similar position are hard to represent the region of attention. Some points are regarded as saccades and discarded, but some small groups remain.

DOI: 10.2352/ISSN.2470-1173.2016.2.VIPC-230

Table 2: The results of different threshold on fixation identification

Threshold	Groups	Percent	Deviation
5	239	0.2166	3.741
10	310	0.5823	6.9992
15	175	0.8078	8.5612
20	99	0.9126	13.0698
25	49	0.9602	13.6494
30	32	0.9756	27.3289
35	25	0.9814	27.3289

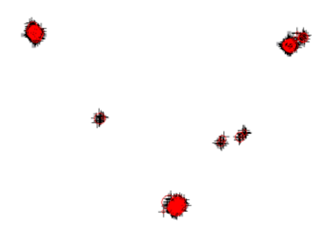


Figure 6*. The result of fixation identification with the threshold of 15. The red circles represent the groups generated after the process.*

Experiments on clustering of ROAs

We utilize the result from the former step to perform the experiments. After the process of the identification, we get 175 clusters. We apply different threshold on clustering of ROAs, and evaluate the result according to the number of cluster and the maximum variance of all cluster. As the threshold increases, the number of clusters decreases and the variance increases. But appropriate threshold will not produce a cluster with high variance. Because most saccades have been discarded, the merged clusters can represent the ROAs properly.

Table 3: The results of different threshold on clustering of ROAs

Threshold	Groups	Percent	Deviation
20	5	0.7185	10.4574
25	6	0.7564	10.4574
30	5	0.7731	10.7574
35	3	0.7731	10.6281
40	3	0.7847	10.6281
45	3	0.7847	10.6281

Figure 7*. With the threshold of 35, large number of fixation are clustered into 3 region-of-attentions.*

Experiments on locating densest center

After the former process, we obtain the clusters direct toward ROAs. In each cluster, we perform the proposed method to locate the densest center. Figure 8 shows the densest center of each cluster, which can well locate the point where tester is interested at. In Figure. 9, the center of the crosshair acts as the ground truth. The mean based method and the densest based method have the deviation of 5.5 and 1.8 pixels from the ground truth separately.

The I-VT algorithm is easily affected by the noise near the fixation point and the velocity threshold is not easy to control. Therefore, the generated clusters may contain some errors. The mean based method for the center of the cluster are very prone to these errors. By contrast, the method we proposed has a better performance in locating the center that we are interested at. Compare to the traditional I-VT algorithm, the proposed method performs superior in separation of fixation and locating the interested center.

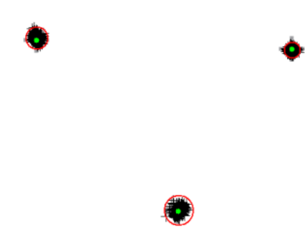


Figure 8*. Process with the proposed method, the green points are the densest center of the cluster. It accurately locates the interested center of the region-of-attention.*

DOI: 10.2352/ISSN.2470-1173.2016.2.VIPC-230

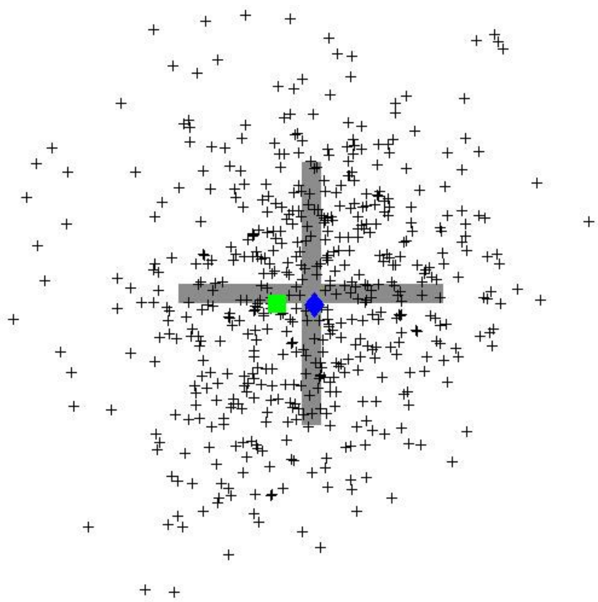


***Figure 8**. The participants are instructed to focus on the center of the crosshair (gray) which is regarded as the ground truth. The center using mean based method (green square), and densest based method (blue diamond) show the deviation of 5.5 and 1.77 pixels, respectively.*

Conclusion

In this paper, we propose a novel method for the identification of visual attention. We use velocity-threshold based fixation identification algorithm to identify fixations and divide them into different clusters. For each cluster, we propose a densest position based method to calculate the center of ROA. As demonstrated in our results, the proposed method shows superior performance when compared to the traditional method.

Acknowledgement

This work was supported in part by National Natural Science Foundation of China (No. 61471273, 91438203), National High-tech R&D Program of China (863 Program, 2015AA015903), and Natural Science Foundation of Hubei Province of China (No. 2015CFA053).

References

[1] A. Duchowski, Eye tracking methodology: Theory and practice, Springer Science & Business Media, 2007.

[2] S. Martinez-Conde, S. L. Macknik, and D. H. Hubel, "The role of fixational eye movements in visual perception," Nature Reviews Neuroscience, vol. 5, no. 3, pp. 229–240, 2004.

[3] B. Cassin and S. Solomon, "Dictionary of eye terminology," 1990.

[4] D. D. Salvucci, "Mapping eye movements to cognitive processes," Ph.D. dissertation, Carnegie Mellon University, 1999.

[5] A. T. Duchowski, "A breadth-first survey of eye-tracking applications," Behavior Research Methods, Instruments, & Computers, vol. 34, no. 4, pp. 455–470, 2002.

[6] T. Busjahn, C. Schulte, B. Sharif, A. Begel, M. Hansen, R. Bednarik, P. Orlov, P. Ihantola, G. Shchekotova, M. Antropova et al., "Eye tracking in computing education," in Proceedings of the tenth annual conference on International computing education research. ACM, 2014, pp. 3–10.

[7] R. A. Hurley, A. Ouzts, J. Fischer, and T. Gomes, "Effects of private and public label packaging on consumer purchase patterns," Packaging Technology and Science, vol. 26, no. 7, pp. 399–412, 2013.

[8] A. Poole and L. J. Ball, "Eye tracking in hci and usability research," Encyclopedia of human computer interaction, vol. 1, pp. 211–219, 2006.

[9] C. J. Erkelens and I. M. Vogels, "The initial direction and landing position of saccades," Studies in Visual Information Processing, vol. 6, pp. 133–144, 1995.

[10] P. Blignaut, "Fixation identification: The optimum threshold for a dispersion algorithm," Attention, Perception, & Psychophysics, vol. 71, no. 4, pp. 881–895, 2009.

[11] R. Den Buurman, T. Roersema, and J. F. Gerrissen, "Eye movements and the perceptual span in reading," Reading Research Quarterly, pp. 227–235, 1981.

[12] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in Proceedings of the 2000 symposium on Eye tracking research & applications. ACM, 2000, pp. 71–78.

[13] H. Widdel, "Operational problems in analysing eye movements," Advances in Psychology, vol. 22, pp. 21–29, 1984.

[14] J. H. Goldberg and J. C. Schryver, "Eye-gaze determination of user intent at the computer interface," Studies in Visual Information Processing, vol. 6, pp. 491–502, 1995.

[15] C. M. Privitera and L. W. Stark, "Algorithms for defining visual regions-of-interest: Comparison with eye fixations," IEEE Transactions on Pattern Analysis and Machine Intelligence , vol. 22, no. 9, pp. 970–982, 2000.

[16] T. Urruty, S. Lew, N. Ihadaddene, and D. A. Simovici, "Detecting eye fixations by projection clustering," ACM Transactions on Multimedia Computing, Communications, and Applications, vol. 3, no. 4, p. 5, 2007.

[17] Guestrin, Elias Daniel, and Moshe Eizenman. "General theory of remote gaze estimation using the pupil center and corneal reflections." IEEE Transactions on Biomedical Engineering, 53.6 (2006): 1124-1133.