

Revisiting Visual Attention Identification Based on Eye Tracking Data Analytics

Yingxue Zhang ¹, Zhenzhong Chen ²

School of Remote Sensing and Information Engineering, Wuhan University

Wuhan, Hubei, China

¹ grace@whu.edu.cn

² zzchen@whu.edu.cn

Abstract—Visual attention identification is crucial to human visual perception analysis and relevant applications. In this paper, we propose a comprehensive visual attention identification algorithm consisting of clustering and center identification. In the clustering process, a spatial-temporal affinity propagation method for accurate fixation clustering is proposed. For the identified clusters, the random walk based method is utilized to extract the center of each cluster, which presents the essential part of an area of interest (AOI). The proposed approach addresses the problem of fixation overlapping in eye movement analytics. Compared with the state-of-the-arts methods, the proposed method shows superior performance for different eye tracking experiments.

Index Terms—Visual attention; eye tracking; clustering; affinity propagation; random walk

I. INTRODUCTION

Eye movements on visual targets attach much importance to cognition-related researches [1]. The eye tracking data can quantitatively reflect the visual perception behaviors. In related domains, attention is typically paid on eye movements in terms of fixations and saccades.

To apply the raw data to further analysis, we should at first figure out the fixations and saccades. Thus the classification and clustering process should be implemented. Many articles have illustrated applicable algorithms for classification and clustering, such as Velocity-Threshold Identification(I-VT), Hidden Markov Model fixation Identification (I-HMM) [2], Dispersion-Threshold Identification (I-DT) [3], K-means [4], projection techniques and density-based combing clustering [5], and agglomerative hierarchical clustering [6]. The interpretation of eye movements varies greatly when different algorithms or parameter settings are applied.

The fixation centers make a crucial indicator of how the subjects comprehend different objects. While viewing a target, the subjects tend to be attracted by salient objects that can be simplified as centers of AOIs, so that the fixations mostly gather around certain centers [7]. With fixation centers identified, further applications that need accurate visual focus, such as eye tracking assisted human-computer interaction, virtual

reality, can be guaranteed. In most situations, the mean based methods are applied to generate centers of visual attention [8].

However, there are still some problems existing in the visual attention identification process. Most clustering algorithms adopted in the current eye tracking systems are conducted merely in single dimension, leading to either redundant results or lack of practical meaning. Moreover, for visual attention center identification, the widely-used mean based methods ignore the inner spatial relation among fixations by taking each point equally and are sensitive to noises. Under this circumstance, the fixation cluster overlapping and center deviation are critical problems in eye movement analytics.

Considering the deficiency, we propose a visual attention identification algorithm based on the combination of spatial-temporal affinity propagation clustering and random walk. The algorithm takes account of varieties of attributes in eye movement data, such as distance, duration and density, to handle the problem and proves improved performance in the experiments.

The paper is arranged as follows. Section II illustrates our comprehensive analytics of eye tracking data for visual attention identification. Section III shows the experiments. Section IV makes a conclusion for the paper.

II. VISUAL ATTENTION IDENTIFICATION

Given a set of raw eye tracking data, the I-VT algorithm is implemented firstly to remove saccades and obtain the initial fixation clusters separated by saccades. Then random walk is conducted on each cluster to generate initial centers which will be clustered with the affinity propagation. The extra clustering process ensures the centers around the same AOI being merged. The initial clusters will also be merged to form the final clusters accordingly using affinity propagation. For each final cluster, we perform random walk based method again to identify final centers of AOIs. The overall work flow is shown in Fig. 1.

A. Clustering Eye Tracking Data

In this work, the spatial-temporal clustering is conducted on the basis of I-VT and affinity propagation. I-VT is widely used in the eye tracking data classification and clustering process for its simplicity. However, since only velocity is considered,

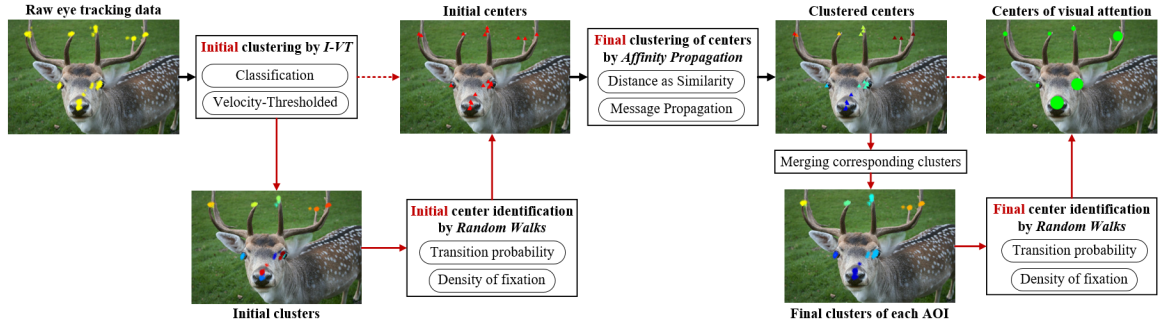


Fig. 1. The work flow of the proposed method.

the algorithm faces the problem of cluster overlapping in most cases, which is the primary problem we aim to solve.

When the subjects focus on a stimulus, groups of points will be recorded, which may include some spatially overlapped groups. If the overlapped groups are close enough, they are considered to belong to one AOI being reviewed several times. But I-VT can not figure out the relationship for it merely gathers temporally consecutive fixations separated by velocity. In this case, there will be redundant clusters around one AOI, which makes an interference to further analysis.

Concerning the problem, we combine affinity propagation [9] with I-VT. Affinity propagation is a robust algorithm in the spatial dimension. It takes spatial distance as similarity so that it ideally handles the problem of overlapping in I-VT.

1) *Classification and Initial Clustering*: Since the saccades move at a much higher velocity than fixations, I-VT separates the fixations from saccades, discards the saccades and gathers consecutive fixations into clusters with a velocity threshold [2]. By this means, we obtain initial fixation clusters reflecting the moving trajectory of visual attention. The velocity of an eye movement point can be equivalent to the Euclidean distance between the point and the next in calculation on account of a constant recording rate. The velocity threshold is set to 20 for a better final consequence.

2) *Initial Center Identification with Random Walks*: After I-VT process, initial clusters are generated. Provided that the revisit condition exists, there should be extra clusters that must be merged. The merging criteria is the distance between two clusters. To appropriately calculate the distance, we implement random walk on each initial cluster to identify a center which can best represent the cluster in distance calculation. Since the random walk based method utilized here is the same as the final center identification process, the specific algorithm details will be introduced in the center identification section.

3) *Final Clustering Using Affinity Propagation*: Obtaining the centers representing the initial clusters, we conduct affinity propagation on the centers in order to merge the clusters belonging to the same AOI. We use the centers to participate in the final clustering for the centers identified by random walk can best represent the spatial position of the clusters.

- *Establishing similarity matrix*: The similarity among points is the basis of spatial clustering. A similarity

matrix among all the initial centers is established for clustering. The similarity $s(i, j)$ is defined using the negative Euclidean distance between point i and j . To generate a moderate cluster number, we set the preference, i.e. self-similarity, to half of the median of all the similarities.

- *Message propagating*: Two kinds of messages representing the affinity, i.e. “responsibility” and “availability”, are defined and recursively propagated till refined clusters emerge. The availability is initialized to zero while the responsibility is initialized and updated as:

$$r(i, k) = s(i, k) - \max_{k' \neq k} \{s(i, k')\}, \quad (1)$$

where k' means the other candidate centers except k . The availability is updated by:

$$a(i, k) = \min \left\{ 0, r(k, k) + \sum_{i' \notin \{i, k\}} \max \{0, r(i', k)\} \right\}, \quad (2)$$

where i' means the other candidate centers except i, k . Specially, the self-availability is updated differently as:

$$a(k, k) = \sum_{i' \neq k} \max \{0, r(i', k)\}. \quad (3)$$

- *Damping factor*: To avoid numerical oscillations arising in unexpected circumstance, we add a damping factor to the messages in every iteration:

$$r(i, k) = (1 - \lambda) r(i, k) + \lambda r_{old}(i, k), \quad (4)$$

$$a(i, k) = (1 - \lambda) a(i, k) + \lambda a_{old}(i, k). \quad (5)$$

where λ is the damping factor between 0 and 1. We set λ to 0.9 in this paper. $r_{old}(i, k)$ and $a_{old}(i, k)$ are the messages in the previous iteration.

- *Identifying final clusters*: When the message propagation accomplishes, the convergent matrices of r and a are added together to form an evidence matrix E . We extract the diagonal elements of E to determine the clustering result. The point k whose corresponding $E(k, k) > 0$ will be chosen as an exemplar. Meanwhile, non-exemplar points will be assigned to the cluster centralized with the exemplar that has the largest similarity with them.

While the centers belonging to the same AOI are clustered into one group, the initial fixation clusters represented by the centers will also be merged correspondingly.

B. Identifying Visual Attention Centers

From the clustering process, eye fixations are divided into clusters around different AOIs. On each fixation cluster, we conduct random walk again to figure out its final center, which represents the visual focus. The identification is premised on the assumption that a center is surrounded by a large percentage of fixations of high consistency with each other [10]. Random walk assigns a coefficient to each fixation according to its potential approximation and transmits it to the neighbors given a high consistency. Compared with mean or density based methods in single dimension, random walk combines both spatial and temporal cues to locate the final centers.

- *Defining transition probability:* The transition probability $q(i, j)$ from point i to j is calculated using the equation:

$$q(i, j) = \frac{e^{-\sigma \times D(i, j)}}{\sum_{k=1}^n e^{-\sigma \times D(i, k)}}, \quad (6)$$

where $D(i, j)$ is the Euclidean distance from i to j . σ makes a subtle adjustment to the distribution of centers and the denominator normalizes the probability. σ is set to 0.08 here. The transition probability reflects the approximate probability between every two fixations. A farther distance leads to a smaller approximate probability and vice versa.

- *Integrating fixation density:* For each fixation, its coefficient is initialized using the density of relevant fixations, which is integrated on the basis of tracking duration. The coefficient is obtained by normalizing the density.
- *Updating coefficients with random walk:* Random walk recursively updates the coefficients using the transition probability of fixations. To reduce the input errors, a damping factor is added to the process:

$$l_{t+1}(i) = \frac{1}{\eta} \left(\sum_{j=1}^n (1 - (1 - \alpha)l_t(i))l_t(j)q(j, i) + (1 - \alpha)l_t(i)w(i) \right), \quad (7)$$

where $l_t(i)$ means the coefficient of fixation i in iteration t . The damping factor is expressed as $(1 - \alpha)l_t(i)w(i)$. α is set to 0.5. η is the parameter that normalizes the coefficient.

$$\eta = \sum_{i=1}^n \left(\sum_{j=1}^n (1 - (1 - \alpha)l_t(i))l_t(j)q(j, i) + (1 - \alpha)l_t(i)w(i) \right). \quad (8)$$

Iteration terminates while reaching convergence that the coefficient l_{t+1} is equal to that of the previous iteration.

- *Identifying fixation centers:* Finally, We obtain the center $(\hat{x}, \hat{y})$ of a certain fixation cluster by calculating the mean fixation position weighted by the final coefficient l_T , which is obtained from the updating process.

$$\begin{cases} \hat{x} = \sum_{i=1}^n x_i l_T(i) \\ \hat{y} = \sum_{i=1}^n y_i l_T(i) \end{cases} \quad (9)$$

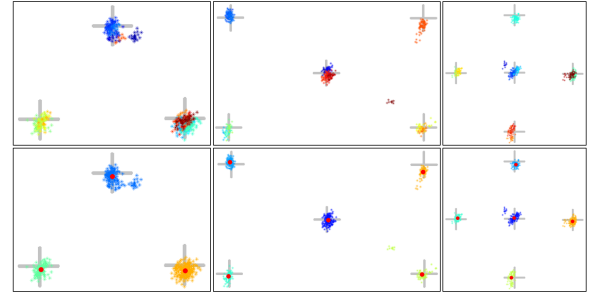


Fig. 2. Experimental results on patterns. The first row is the result of initial clustering with I-VT. The second row is the final result of our method. Fixations of different clusters are marked out with asterisks of different colors. The centers are represented with red dots. The grey crosses show the ground truth.

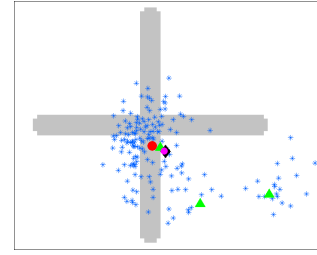


Fig. 3. The magnified partial view of the result in Fig. 2. The identified centers are marked out with red dot (our method), green triangle (Tobii's default method), black diamond (K-means) and magenta dot ([11]).

III. EXPERIMENTS

A. Experiment Setup

We collect the eye tracking data with a Tobii X120 Eye Tracker. The tracker is at a distance of 1 meter from the subject, tilt for 30 degrees and placed in front of a 27-inch computer monitor that presents the stimuli. Each stimulus is viewed for about 10 seconds. The results, including coordinate, duration and recording moment of each eye movement point, are recorded at 120Hz. To comprehensively verify the method, we set two experimental scenes and also compare the consequences with some existing algorithms.

B. Experiments on Different Patterns

Three patterns are used for validation, in which the centers are marked out as ground truth. The subject is asked to fix attention on the centers of the patterns. As is shown in Fig. 2, the proposed algorithm avoids the interference of the cluster overlapping and obtains reasonable results.

TABLE I
COMPARISON OF ABSOLUTE PIXEL DEVIATION OF DIFFERENT METHODS.

	K-means based method	Tobii's default method	Špakov's method [11]	Our method
Pixel deviation	13.0822	10.8432	12.6322	8.4430



Fig. 4. Experiment on natural images. The results of different methods are shown in corresponding columns.

We also conduct other methods including K-means based method, density based method in [11], and Tobii's default method utilized in Tobii eye tracker software. In Fig. 3, the results of these methods on a single center are shown to illustrate the advantage of our method. We can see that our algorithm locates the center with the smallest deviation. For the Tobii's method, two mistaken centers are marked out on account of merely considering temporal factor. The results of K-means and density based methods distinctly deviate from the ground truth. Table I lists the absolute pixel deviation from the ground truth in the partial view, which quantitatively shows the advantage of the proposed method.

C. Experiments on Natural Images

To validate the proposed method in practical eye-tracking applications, we implement the algorithms on natural images from [12]. We can see from Fig. 4 that the proposed method successfully handles the messy overlapping fixations and identifies one center on each AOI, while Tobii's default method faces failure of generating too many unnecessary centers. In comparison, the results of density based method and K-means based method are similar to ours. However, the density based method is easily influenced by the outliers around the AOIs and deviates towards the outliers. K-means based method is quite sensitive to the initialization so that it has to be rerun at least three times to get an acceptable result and the number of clusters has to be set manually for each image.

IV. CONCLUSION

In this paper, we present a comprehensive algorithm of visual attention identification, which is based on clustering and visual attention center identification. On the eye tracking dataset, fixation clusters are generated with our proposed spatial-temporal affinity propagation clustering method. On each cluster, random walk based method is conducted to identify the corresponding center. The algorithm solves the problem of overlapping in eye movement analytics and achieves a more accurate center identification result. In the experiments, we verify the effective performance of our algorithm with

two experiments. In comparison with other methods, the proposed visual attention identification algorithm outperforms in accuracy and robustness.

ACKNOWLEDGMENT

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).

REFERENCES

- [1] L. Mason, P. Pluchino, and M. C. Tornatora, "Eye-movement modeling of integrative reading of an illustrated text: Effects on processing and learning," *Contemporary Educational Psychology*, vol. 41, pp. 172 – 187, 2015.
- [2] 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.
- [3] 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.
- [4] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," *Pattern Recognition*, vol. 36, pp. 451–461, 2003.
- [5] T. Urruty, S. Lew, C. Djeraba, and D. A. Simovici, "Detecting eye fixations by projection clustering," *ACM Transactions on Multimedia Computing Communications and Applications*, vol. 3, pp. 1–20, 2007.
- [6] A. Bouguettaya, Q. Yu, X. Liu, X. Zhou, and A. Song, "Efficient agglomerative hierarchical clustering," *Expert Systems with Applications*, vol. 42, pp. 2785–2797, 2015.
- [7] C. Privitera and L. 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.
- [8] S. D. Krönig and E. A. Buffalo, "A nonparametric method for detecting fixations and saccades using cluster analysis: Removing the need for arbitrary thresholds," *Journal of Neuroscience Methods*, vol. 227, pp. 121–131, 2014.
- [9] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, pp. 972–976, 2007.
- [10] A. R. Zamir, S. Ardeshtir, and M. Shah, "Gps-tag refinement using random walks with an adaptive damping factor," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2014, pp. 4280–4287.
- [11] O. Špakov and D. Miniotos, "Application of clustering algorithms in eye gaze visualizations," *Information Technology and Control*, vol. 36, no. 2, pp. 213–216, 2007.
- [12] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *IEEE 12th International Conference on Computer Vision*, 2009, pp. 2106–2113.