A Novel Method for Scanpath Comparison Based on Levenshtein Distance

Xiang LI and Jiacai ZHANG
College of Information Science and Technology, Beijing Normal University, Beijing, China
Email: lixiang_cn@foxmail.com, jiacai.zhang@bnu.edu.cn

***Abstract*—Modeling the scanpath of human saccade on presentation pictures is a challenging task. To evaluate the consistence of saccadic eye movement sequences, traditional methods, such as scanMatch, transform scanpath data to string sequence with grid-based method and computer the similarity of string sequences using Needleman-Wunsch algorithm. However, scanpath data are different from string data in some aspects, and grid-based method would cause errors in measurement. In the present work, the difference between two symbol sequences was measure by a modified Levenshtein Distance (LD) method. Here, the pair-point distance was computed from position of different fixation points directly, and the distance was further converted to 0 or 1 by a thresholding method. In this way we can adjust the matching threshold value between points and improved the accuracy of LD method by overcoming measurement error of point around grid boundary. Our experimental results demonstrated that our method had better performance than traditional scanMatch.**

***Index Terms*—scanMatch, levenshtein Distance, fixation, scanpath similarity**

I. INTRODUCTION

The attention mechanism of human visual system can extract the most important parts from a large amount of redundant visual input for the further processing of cerebral cortex [1]. In the past decades, the research on the attention mechanism has become a hot topic [2]-[4]. Recently researches on human attention transfer are mainly focused on fixation points and saccades [5]-[6]. Saccades indicate rapid eye movements the eyes move quickly from one position to another. The eye is stationary between saccades, the period between saccades is defined as fixation, and useful visual information can be collected by human visual systems during the time intervals [7]. And the scanpath, the sequences of saccadic eye movements, consists of position, time and order information.

In the most current studies on eye movements, fixation point distribution [8], fixation duration [9], saccade amplitude [10], or other information have been used as the main measures. However, there are few available methods to investigate the sequential properties between fixation points. Some scholars have used hidden Markov process to model scanpath and the factors that may influence scanpaths were explored [6], [11]. However, the implementation process is complicated and it is difficult to quantify the differences between the two scanpaths. String matching plays an important role on natural language processing (NLP) field. Editing distance, also called Levenshtein Distance (LD), as a method to measure the difference between strings, has been widely applied and developed [12]. Similarly, it has been applied to measure the difference between different scanpaths [13]. However, due to the inconsistency between scanpaths and strings, there are some problems in measuring the difference of scanpaths by editing distance. For example, the similarity between fixation points is ignored, resulting in a large measurement deviation at some time.

In the field of biogenetics, DNA alignment is also a hot topic for research on temporal information. Needleman-Wunsch algorithm, based on the implementation of dynamic programming, solves the alignment problem and performs well in DNA alignment [14]. The basic idea is creating the best overall alignment by the local optimal alignment of sub-sequences. By doing this, it reduces the number of possibilities to be considered but still guarantees the optimal solution. The algorithm is divided into two parts; the first creates a matrix with all scoring possibilities based on a substitution matrix and a gap penalty, and the second seeks the optimal alignment, tracing back the matrix from the top left corner to the outmost column or row by selecting the optimal route. The score is given by the optimal route throughout the matrix; the higher the score is, the more likely it is that the two DNA sequences are similar.

ScanMatch method is a successful application of Needleman-Wunsch algorithm in the study of eye movement trajectory. It not only considers the location and time information, but also measures the order information, and has achieved good results [15]. The method is divided into two parts. First, the scanpath was converted into coding sequences so that Needleman-Wunsch algorithm can be applied on scanpath. Specifically, the image is segment into different subregions by the defined regions of interest (ROI) on the image, and each subregion is assigned a unique encoding. Then, the fixation point is represented by the corresponding code in the region where it falls, and the code is repeated according to the duration of fixation point, and finally the corresponding coding sequence of

Manuscript received December 25, 2018; revised May 10, 2019.

doi: 10.18178/ijsps.7.2.32-37

scanpath is obtained. In the second part, Needleman-Wunsch algorithm is applied to calculate the similarity of coding sequences as the similarity of scanpath based on the substitution matrix and gap penalty. ScanMatch takes into account the similarity of features between fixation points when designing the replacement matrix, which is ignored in the editing distance, and successfully solves some problems existing in the editing distance. However, just like the editing distance, scanMatch is a grid-based method, so it is inevitable to encounter problems caused by gridding.

Several criticisms have been raised against the grid-based method for scanpath comparisons (Fig. 1). Firstly, two fixations may be considered different even when they are close together, namely even if they fall on the either side of a grid line [16]. Secondly, all fixation points in a region are considered to be the same, which means that the difference between the two most distant fixation points in two adjacent regions is the same as that between the two closest ones (Fig. 1). Moreover, although regions of interest are artificially identified, the definition of ROI is difficult and time-consuming, and it is difficult to ensure that there is an appropriate region division scheme for each image in the actual scene.

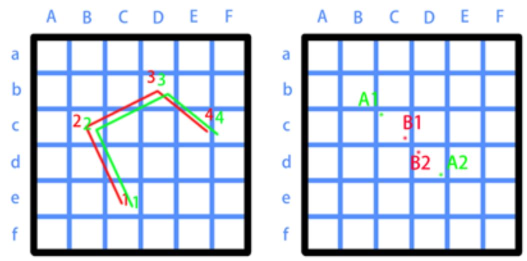

Figure 1. Problems of grid-based method. In the left figure, the fourth points in two scanpaths are similar, but they are divided into different regions. In the right figure, the difference between A1 and A2 is the same as the difference between B1 and B2.

Based on this, we improved the traditional editing distance method and proposed scanLD method to measure the similarity of scanpaths. This method takes into account the time and order information, and at the same time describes the relationship between the cost of editing distance and the position of fixation points through a function. By doing so, we fused the position of fixations into editing distance cost, so as to obtain more accurate similarity measurement between scanpaths and solve the measurement error problems caused by grid-based methods mentioned above. Through experiments, it proved that the proposed method could measure the similarity between scanpaths more robustly and accurately. We implement it on the python platform, and all the code is available on GitHub (https://github.com/DreamLee0625/scanLD).

II. Materials and Methods

A. Data Pre-processing

Generally, scanpath data included a series of fixation points arranged in sequence, and each fixation point included position and duration information. In order to speed up the calculation on the basis of keeping the time and order information, we first resampled the original data output by the eye tracker in temporal dimension, so the amount of data can be reduced. Specifically, the position data of fixation points are repeated proportionally according to the fixation duration, and the sequence data of eye movement trajectory to be compared is obtained (Fig. 2).

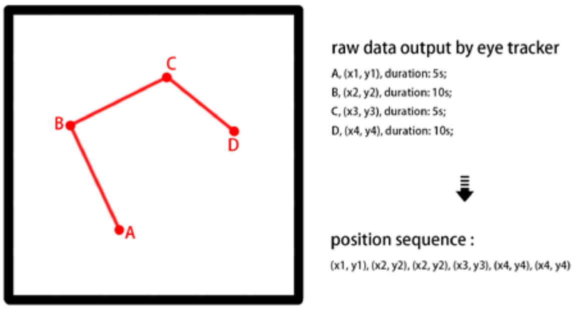

Figure 2. Data pre-processing.

B. Scanpaths Comparison

The purpose of the method presented in this article is to quantitatively characterize the similarity between two scanpaths are (i.e., “(100, 100), (200, 200), (300, 300), (400, 400)” with “(100, 100), (200, 200), (300, 300), (400, 300)” or “(100, 100), (200, 200), (240, 240), (400, 400)”). Therefore, instead of grid-based methods, we improved the traditional editing distance to make it more robust and accurate to measure the similarity between scanpaths.

C. Edit Distance

Edit distance, also known as Levenshtein Distance, is used to quantify similarities between strings in the NLP field. Edit distance defines three edit operations: insertion, deletion, and substitution. Through several editing operations, one string can be converted to another string. The least number of total editor operations one string can be converted to another string is used as the minimum editing distance. Finally, according to the minimum edit distance, the normalized similarity score is obtained by the following formula (1). The minimum edit distance is implemented by dynamic programming. The basic idea is that the global optimal solution can be obtained by solving the local optimal solution, just like Needleman-Wunsch algorithm. However, the calculation of edit distance does not need to trace back, so the edit distance method is faster. The existing methods measuring scanpath with LD ignored relationship between elements. Previous studies converted the scanpath into coding sequence by segmenting the image into regions, and used the edit distance to measure the similarity of scanpath, and achieved certain results [13]. Nevertheless, grid-based methods caused additional measurement errors in the conversion process. Here, the operation segment the image into regions wasn’t adopted, we modified the elements matching strategy and added fixation point feature information into the editing operation, so that the editing distance can better measure the similarity of scanpaths. The matching strategy between elements was introduced in Section D.

$$score = 1 - \frac{editDist}{\max(length(str1), length(str2))} \tag{1}$$

D. Element Matching Strategy

Unlike the fixed division of the ROI on the image, we proposed a hyper-parameter called *match-range* which means the matching threshold of distance. In the process of calculating the edit distance, we dynamically calculate the distance between each pair of fixation points. When the distance between two fixation points was less than the *match-range*, we considered them to be matched. Otherwise, the editing action was needed. By doing this, the problem of misjudgment of fixation point near the boundary is solved to some extent. Moreover, if the two fixations are close enough, they are matched. However, there still exists some situations where the distance between the two points is just above the threshold. This problem will be solved in next section by adding distance to editing operation cost.

E. Cost of Editing Operations

In this paper, we took the distance between the two fixations into consideration when calculating editing operation cost. By doing this, LD can measure the difference between two scanpaths better. Specifically, for editors cost of insertion and deletion operation, the operation of the conversion between two states were described as "one" and "zero", so their editing cost should be the biggest, is 1. The editing cost of the substitution operation is determined by the similarity between the two elements. The more similar the two elements are, the smaller the editing cost will be; the greater the difference between the elements, the greater the editing cost will be. However, the editing cost of the substitution operation will not exceed 1. This is because if the editing cost of the substitution operation is greater than 1, the insertion and deletion operation with less editing cost will be taken instead of the substitution operation at each step of calculating the edit distance. Based on the above ideas, we proposed a function to describe the relationship between the distance and the cost of substitution operation. The function input is the distance between fixation points, and the function output is the cost of substitution operation, and the output value should be between 0 and 1. Here, we simply use piecewise functions, as Fig. 3.

Match-range is the boundary of the minimum cost of the substitution operation, while *inception-range* is the boundary of the maximum cost of the substitution operation (Fig. 3). When the distance between fixation points is greater than the maximum boundary, they are no longer related to each other, so the editing cost of the substitution operation between them is 1. In this way, improved LD for eye movement trajectory can be obtained by the following threshold function (2):

$$editCost = \begin{cases} 0 & dist(p_1, p_2) <= matchRange \\ \min\begin{cases} insertCost \\ substitutionCost \\ deleteCost \end{cases} & dist(p_1, p_2) > matchRange \end{cases}$$

$$\begin{aligned} & insertCost = 1 \\ & deleteCost = 1 \\ & substitutionCost = f(dist) \end{aligned} \tag{2}$$

Here, where the function f is describes the relationship between distance and editing cost, in this article, it's described as formula (3):

$$f(dist) = \frac{dist - matchRange}{inceptionRange - matchRange} \tag{3}$$

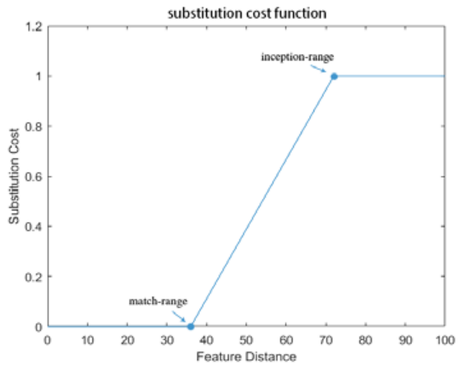

Figure 3. Substitution Cost Function.

F. The Scoring

The improved LD based on eye movement data is called scanLD. Finally, the similarity score between the two scanpaths is determined by the scanLD method and the length of the longest common subsequence. The formula is as (4):

$$score = \frac{adjustEditCost + longestCommonSubseqs}{longestCommonSubseqs} \tag{4}$$

In this normalization, the two best matches will score 1 and the two worst matches will score as close as possible to 0.

G. Python Code

The implementation of the experimental code is based on python, and more detailed instructions can be found on GitHub (https://github.com/DreamLee0625/scanLD).

III. Experiment and Results

A. Experiment on Simulation Data

In order to demonstrate the method we proposed can overcome boundary point problem caused by grid-based method and the robustness of the method gets much better, the experiment was built on simulation data. Moreover, the comparison on the results of edit distance method and scanMatch method was done.

We set eight sampling center points on the image of 1024×768, and surrounded them into a flat ring. Ignoring point L or point R, the remaining center points traverse clockwise or counterclockwise to obtain eye movement trajectories under four different conditions. Although these four types of eye movements are highly overlapped in space, and only the traversal order is different, we predicted that this operation would result in four different types of eye movement sequences. The way of scanpath generation is shown in the Fig. 4. We use Gaussian sampling to obtain the simulated fixation points of each sampling center, so as to simulate the randomness of each fixation point.

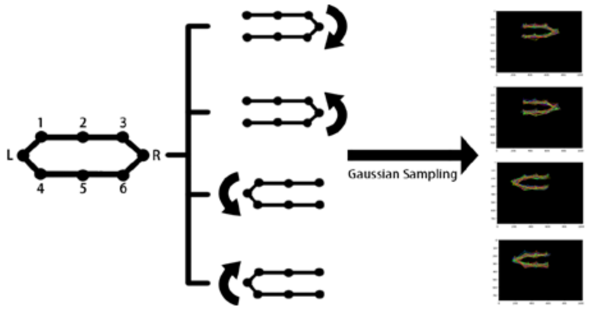

Figure 4. Generation methods of simulation data. Ignoring point L or point R, the remaining center points traverse clockwise or counterclockwise to obtain eye movement trajectories under four different conditions.

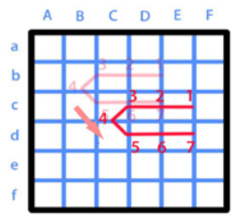

Figure 5. Shifted down of sample center.

In this simple simulation, we assumed a constant fixation time; hence, a single tile represents each simulated fixation. In each experiment, we generated 25 paths for each condition, 100 paths in total. We conducted 9 experiments and gradually shifted the sampling center downward (Fig. 5), and the sampling center will move down 8 times accumulatively, each time it moved down 16 pixels, shown in Fig. 5. For the editing distance and scanMatch method, the eye movements were binned spatially with a grid size of 8*6 according to the demo of scanMatch. For ScanMatch's method, we used the parameters provided from demo. In different experiments, their ROI Settings remained unchanged, so as the sampling center moved down, we expected to show the errors caused by the grid-based method.

In each experiment, we used edit distance, scanMatch and our method called scanLD to compare the similarities between scanpaths. Each scanpath in each experiment was compared with all the rest in the same experiment. Therefore, scanpath belonging to the same conditions should have a high similarity score, while different conditions will have a low similarity score.

A k-mean clustering algorithm was then used to classify each comparison set to either the first set (a within-sets comparison) or the second set (a between-sets comparison).

As shown in Fig. 6. When the sampling center is far from the ROI boundary, the edit distance can distinguish the scanpaths with different conditions better. When the sampling center shifts to the ROI boundary (shift=64), the accuracy of the edit distance method decreases from 100% to 57.69%. As can be seen from Fig. 7, the similarity within class of condition 1 is close to that between different conditions. As can be seen from Fig. 8, the matching matrix between conditions has become much difficult to distinguish different conditions. ScanMatch method, which also belongs to grid-based method, has a fluctuating accurate rate throughout the experiment (Fig. 6) and a high similarity across different classes (Fig. 7 and Fig. 8). It is difficult to distinguish condition 1 from condition 2 (Fig. 7 and Fig. 8), whose spatial distribution is the same, only the traversal order is different. This is because scanMatch takes the position information of the fixation point into account, but it does so based on the area segment. For example, in Fig. 9, the red region is set to be the benchmark in terms of similarity whereas, the upper, lower, left and right region is set to be 0.8, and the oblique direction of the red region is set to be 0.4. Although A is closer to B2, scanMatch shows that the similarity between A, B1 and B2 is the same, which is 0.8. Our method maintains a stable performance in the whole process of center translation. Intra-class similarity is much higher than inter-class similarity. The distinction between across classes is better.

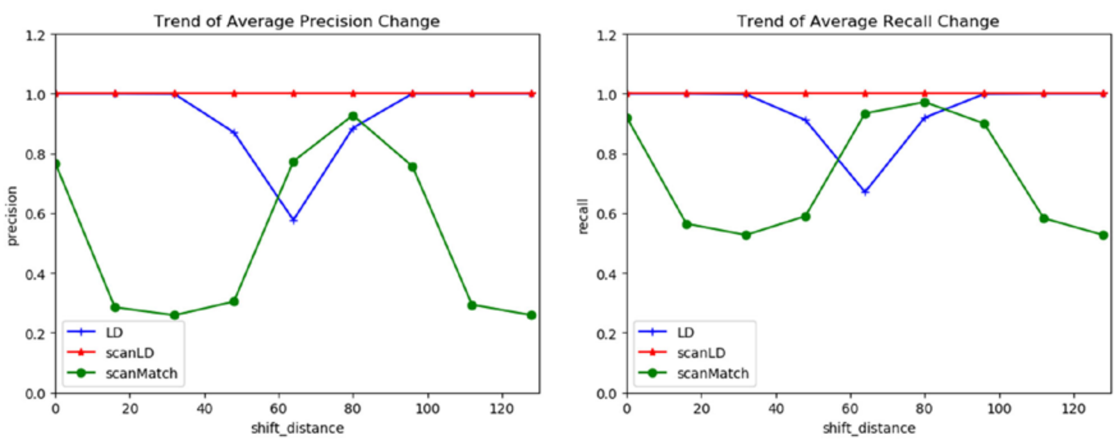

Figure 6. The trend change of precision and recall rates as the sampling center moves down.

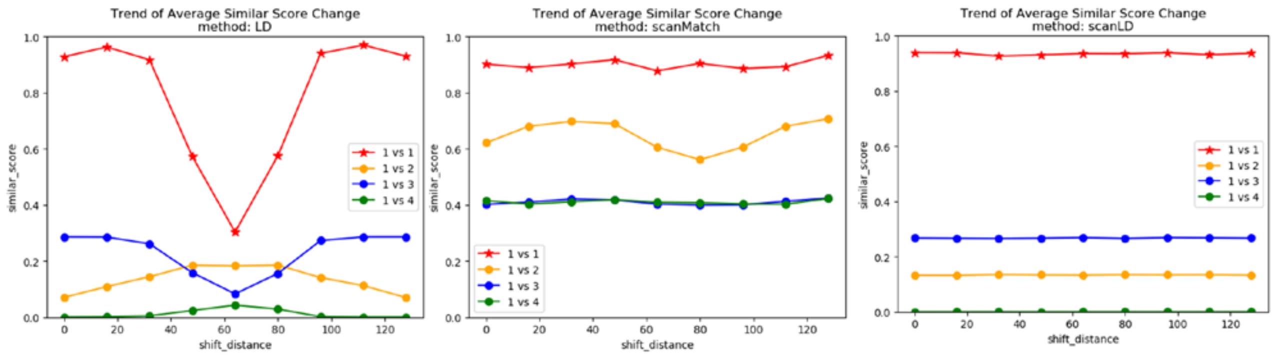

Figure 7. Taking condition 1 as an example, and the similarity between different conditions changes with the moving down of the sampling center.

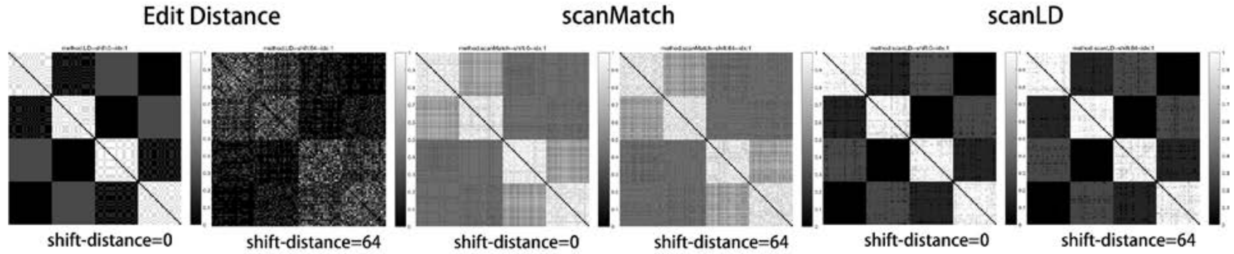

Figure 8. The matching matrix of similarity between different conditions when the sampling center does not move down and when the sampling center moves down to the boundary.

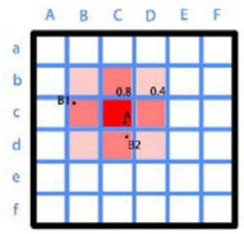

Figure 9. Problem caused by grid-based method.

B. *Experiment on Real Data*

In this experiment, we assessed the performance of the new method, using a simple sequential-looking task with human eye movement data. In this experiment, we presented stimulus images on a computer screen, with green or red numbers randomly placed on the images from 1 to 9. Five subjects in total participated in the experiment. The subjects were asked to look at red or green numbers in ascending or descending order and their eye movements were records. Subjects were given specific instructions before each experiment. To ensure the correctness of the task, each subject will conduct four sets of preliminary experiments to familiarize himself or herself with the experimental process, and then conduct formal experiments. Each command will conduct five sets of experiments, a total of 20 sets. We experimented with Tobii x120 eye tracker, which has a sampling rate of 120 Hz and a maximum scanning Angle of 35 degrees. The stimulus is displayed on a 25-inch display with a resolution of 2560×1440 pixels. Before the experiment starts, the 9-point calibration procedure will be executed and the calibration result will be checked. The extraction of eye movement track is automatically filtered by the ClearView-Fixation-Filter. The stimulus images under different commands are consistent for subsequent eye movement track comparisons. We set the parameters (*match-range* and *inception-range*) both to 64. The temporal re-sampling interval is 50ms. Each scanpath was compared with every other scanpath using our scanLD method. We expect that scanpaths under the same condition will have higher similarity scores than those under different conditions. To check this hypothesis, a k-means algorithm was used to cluster the scoring matrix for each comparison set (a single trial against all of the others). The self-comparison (leading to a perfect score of 1) was removed from the clustering process. The experimental results of each subject are shown in the table I below.

TABLE I. F-RATIO AND P VALUES FOR EACH SCANPATH COMPARISON MEASURE

Subject	Similar score		ANOVA	
	Same-condition	Different-condition	F	p
1	0.83±0.12	0.11±0.04	7021	<.001
2	0.82±0.16	0.10±0.04	4764	<.001
3	0.91±0.10	0.10±0.03	12848	<.001
4	0.86±0.10	0.10±0.05	9284	<.001
5	0.86±0.15	0.12±0.04	5981	<.001

As shown in Table 1, similar score of same-condition trials of all subjects were significantly higher than that of different-conditions. Although the spatial distribution of the same color condition is the same, only the order is different, there is no error in the clustering result. This experiment shows the effectiveness of our method in actual data.

IV. DISCUSSION

In this paper, we propose a new method, called scanLD, to measure the similarity between scanpaths. Our method solves the two problems caused by grid-based method mentioned above. Firstly, this method no longer segments the image by setting the region of interest, but uses the position information of fixation points directly, and proposed the matching judgment method based on the distance, which makes the categorization of two similar points unmistakable. Secondly, the position of fixations point is fused into the editing operation cost. By describing the relationship between distance and editing cost through functions, the improved LD can more accurately describe the differences between fixation points.

Two experiments were carried out to verify the excellent performance of our algorithm. In the first experiment, we simulated the data to generate two problems caused by grid-based method, and compared our method with the editing distance and scanMatch method to illustrate that our algorithm overcame this problem and had better robustness as well as performance. In the second experiment, we collected the data of five subjects, demonstrated the ability of this method to process real data, distinguish scanpaths under different conditions, and have good robustness and accuracy.

The core of our new method is based on Levenshtein Distance algorithm, which has been developed to be computationally efficient without compromising on the quality of the fit between strings. As a result, the comparison of large numbers of strings is possible on a modern desktop computer.

Scanpath is a complex data structure, including position, time and order information, which needs to be considered. Therefore, when measuring scanpath, our method adds its location, time and order information and gets outstanding results. However, our method does not take into account other features of fixation points, such as semantic information around fixation points, which are also important. This hinders our method from measuring semantic similarity of scanpath. Another limitation of our method is that the function used to describe the editing cost and distance is fixed. If the function can be dynamically changed according to the image content, such as the size of the subject in different images, better results may be obtained.

To sum up, our new method, scanLD, can accurately measure the similarity between different eye movements much better. As a measurement method, it provides the possibility for more detailed studies of eye movement trajectory and more accurate analysis of factors affecting eye movement trajectory.

ACKNOWLEDGMENT

This research was funded by the National Key Technologies R&D Program (2017YFB1002502), and the project of Beijing Advanced Education Center for Future Education (BJAICFE2016IR-003).

REFERENCES

[1] L. W. Renninger, J. M. Coughlan, P. Verghese, and J. Malik, “An information maximization model of eye movements,” *Advances in Neural Information Processing Systems*, vol. 17, pp. 1121-1128, 2005.

[2] M. M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S. M. Hu, “Global contrast based salient region detection,” in *Proc. 24th IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 409-416.

[3] C. Yang, L. Zhang, H. Lu, X. Ruan, and M. H. Yang, “Saliency detection via graph-based manifold ranking,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 3166-3173.

[4] Q. Yan, L. Xu, J. Shi, and J. Jia, “Hierarchical saliency detection,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 1155-1162.

[5] W. Zhu, S. Liang, Y. Wei, and J. Sun, “Saliency optimization from robust background detection,” in *Proc. IEEE Conference on Computer Vision & Pattern Recognition*, 2014, pp. 2814-2821.

[6] H. Liu, D. Xu, Q. Huang, W. Li, M. Xu, and S. Lin, “Semantically-based human scanpath estimation with HMMs,” in *Proc. IEEE International Conference on Computer Vision*, 2013, pp. 3232-3239.

[7] R.H.S. Carpenter, *Movements of the Eyes*, 2nd ed. London: Pion Ltd, 1988.

[8] N. C. Anderson, W. F. Bischof, K. E. W. Laidlaw, E. F. Risko, and A. Kingstone, “Recurrence quantification analysis of eye movements,” *Behavior Research Methods*, vol. 45, no. 3, pp. 842-856, 2003.

[9] T. Foulsham, R. Dewhurst, *et al.*, “Comparing scanpaths during scene encoding and recognition: A multi-dimensional approach,” *Journal of Eye Movement Research*, vol. 5, no. 4, pp. 1-14, 2012.

[10] S. V. Shepherd, S. A. Steckenfinger, U. Hasson, A. Ghazanfar, “Human-monkey gaze correlations reveal convergent and divergent patterns of movie viewing,” *Current Biology*, vol. 20, no. 7, pp. 649-656, 2010.

[11] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257-286, 1989.

[12] V. Levenshtein, “Binary codes capable of correcting deletions, insertions, and reversals,” *Soviet Physics–Doklady*, vol. 10, pp. 707-710, 1966.

[13] T. Okuda, E. Tanaka, and T. Kasai, “A method for the correction of garbled words based on the levenshtein metric,” *IEEE Transactions on Computers*, vol. 25, no. 2, pp. 172-178, 1976.

[14] S. B. Needleman, and C. D. Wunsch, “A general method applicable to the search for similarities in the amino acid sequence of two proteins,” *Journal of Molecular Biology*, vol. 48, no. 3, pp. 443-453, 1970.

[15] F. Cristino, S. Mathôt, J. Theeuwes, *et al.*, “ScanMatch: A novel method for comparing fixation sequences,” *Behavior Research Methods*, vol. 42, no. 3, pp. 692-700, 2010.

[16] N. C. Anderson, *et al.*, “A comparison of scanpath comparison methods,” *Behavior Research Methods*, vol. 47, no. 4, pp. 1377-1392, 2015.

Xiang Li is a master student from Beijing Normal University, China. His main research interests include image processing and image recognition, especially on image processing based on eye movement.

Jiacai Zhang received the PhD degree from the Chinese Academy of Science, China. He is a professor and an associate dean at the College of Information Science and Technology of Beijing Normal University. His main research interests include signal processing and image processing.