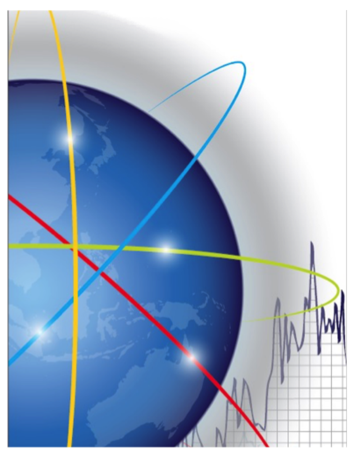

Modelling Eye-gaze Movement by Gaussian Auto-regression Hidden Markov

Beinan Xu[1], **Andy Song**[1]
1. School of Computing Technologies, RMIT University, GPO Box 2476, Australia
Correspondence: xbnxbn38@yahoo.co.jp

Introduction

Predicting eye gaze movement can be very significant in many applications, but the mechanism of human eye gaze movement is very complicated. It is not only related to the information contained in the observed target, but also by different effects of human cognitive behaviors.The main drive in eye gaze movement can be categorized into three groups, (1) the object of interest, which may change quickly due to the complex nature of the scene in front of the observer, (2) interference that often occur in real world scenarios, and (3) uncontrollable movement caused by subconscious reasons yet still to be understood by scientists. Accordingly we aim to separate eye gaze movement into three components, namely the *principle movement* (the movement behavior affected by the primary targets), *interference* and *subconscious movement*.This study mainly focuses on the principle movement in the eye gaze movement, because the interference and subconscious movement are unpredictable and shall be considered as noise and bias, which can be removed from the trajectory data. Principle movement can be modelled by a proposed Gaussian Auto-regression Hidden Markov Model. Finally, the model can predict the probability of occurrence of eye gaze in each region over time. By joining the predicted points together as a sequence, we can generate the eye gaze movement prediction as a time series.

Data and Data pre-processing

Data: The experimental material of our design contains 100+ static images. Also, 20 volunteers were invited. During the experiment, the eye gaze data were collected using 60Hz Tobii II eye tracker, each volunteer is instructed to watch different pictures, each picture appearing for 5 seconds. The total duration of each experiment does not exceed 5 minutes. So we can collect 300 eye movement data points from one volunteer for one image.

Data pre-processing: In order to improve the data quality, K-nearest Neighbor (k=3) and Fourier transform are used to fill the "gaps" in data set, which are generated due to the limitations of the equipment and the personal condition of the subject. The former is to generate missing values while the latter is to adjust the added values so they are consistent with the data distribution.

On the left hand of Figure 1, it is the heatmap of unprocessed eye gaze data, color to signify the intensity of the gaze on a particular spot. There are a few tiny spots far apart from the rest. They are possible outliers in the data set. The right side shows the heatmap generated on the data transformed by K-NN and the Fourier process. Compared to counterparts on the left, it is much smoother and connected. Furthermore the isolated small spots are no longer present. The pre-processing can help extract two dimensional features of the eye movement data more effectively, not only by supplementing values for the missing data points, but also by eliminating the influence of subconscious eye-gaze points.

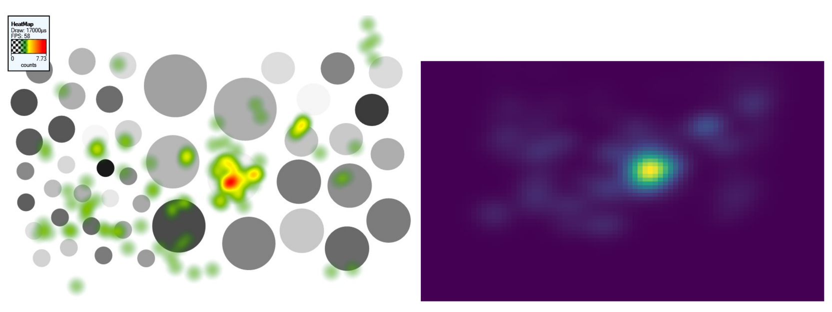

Figure 1: Examples of generated eye gaze heatmap obtained after filling missing data and the Fourier transform.

Methods

Image Transform

This method is one of the core of this research, it can effectively combine eye movement data with image data and extract eye movement behavior. Image information (gray value) contained in a two-dimensional image is quantized to form the image information distribution which is three-dimensional. This initial distribution will constantly change with the eye gaze movement. We divide a target image evenly into a $n \times n$ grid. The larger n is, the viewing behavior on the image, e.g. the eye gaze movement, can be better separated. However that means greater amount of calculation would be needed.

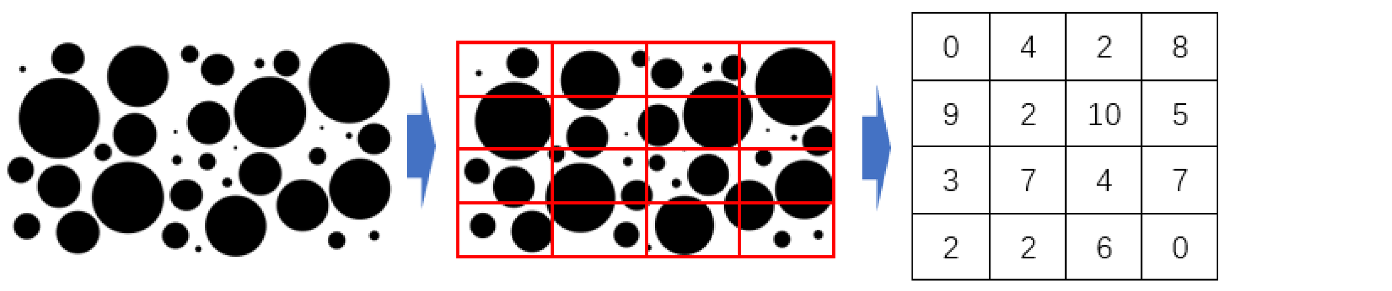

Figure 2: Generating region map based on pixel information. The number of each region represents the cumulative amount level of the total gray scale of the regions (0 to 10), As an example, a target image is marked as 16 smaller regions. Each cell is assigned with a value, which represents the relative cumulative amount level of the total inverted grayscale of that region (white is 0 and black is 255).

Each cumulative gray value in a cell is called the initial influence parameter of eye gaze movement. The values of all regions are normalised to generate an integer value from 0 to 10. It is also the up-bound of changes meaning the value in that cell will never go beyond that point. We can also obtain a value from 0 to 1, which is $1/10$ of the eye gaze influence parameter. This value represents the eye gaze influence coefficient ψ of the image information in the region on the eye movement behavior.

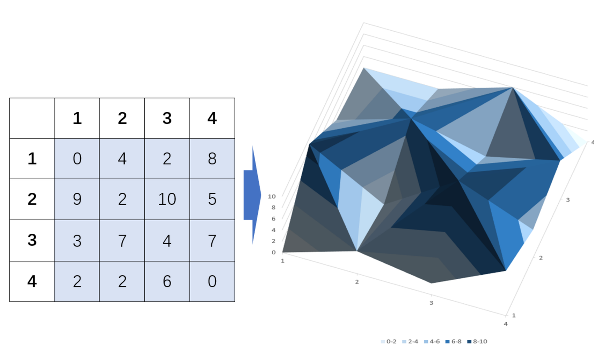

Figure 3: Eye gaze influence parameter distribution 3D-plot base on above example.The shade of the color represents the size of the parameter.

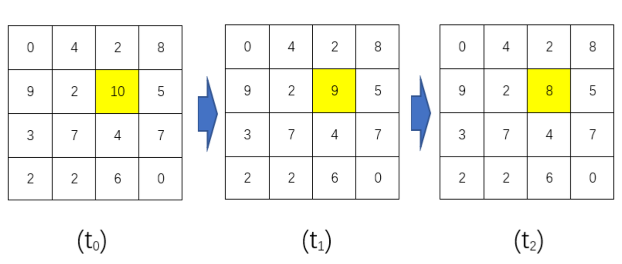

Figure 4: Example 1: When the eye gaze falls on a region, the value of that region will decrease by ψ (10 to 9). If the eye gaze stays in this region from t_0 to t_2, the region will be reduced by $2 \times \psi$ (10 to 8).

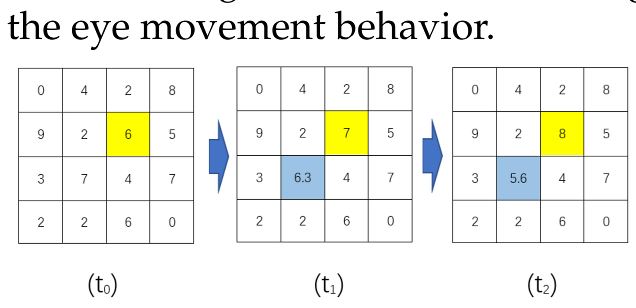

Figure 5: Example 2: The eye gaze falls on the yellow region at time t_0. When the eye gaze moves to the blue region at t_1, the value of yellow region rises with ($\psi = 1$) per interval until its initial value 10. The value of the blue region then starts to decrease with ($\psi = 0.7$) per interval until 0 or eye gaze goes elsewhere.

As the eye gaze moves along, these eye gaze influence parameters will change base on the eye gaze influence coefficient, as illustrated in Figures 4 and 5. If the eye gaze stays in a region, the area influence parameter value will decrease by the coefficient ψ until it reaches zero. If the eye gaze leaves, the parameter value will increase by the coefficient ψ until it reaches the up-bound, that is the initial value.

Gaussian Auto-regression Hidden Markov

Another core of this research, after all changes in eye gaze are updated in these divided regions by Image Transform method, the Gaussian auto-regression hidden Markov model is then applied to capture and analyze these eye gaze movement.

Our proposed GAR HMM is base on AR HMM, Auto-regression hidden Markov with a beta process, which was developed by Emily B. Fox, Michael I. Jordan, etc. Instead of the prior beta process used in AR HMM, we introduce a Gaussian process, which can estimate the probability of one behavior appearing in a particular region rather than a binary decision "appear" or "not-appear" in that region. See the diagram of the GAR HMM model in right figure, where f_i represents the probability within the acceptable range of k. A feature-constrained transition distribution $pi = pi_{z_{t-1}}$ is defined, which controls the Markov transition between its dynamic behavior sets.

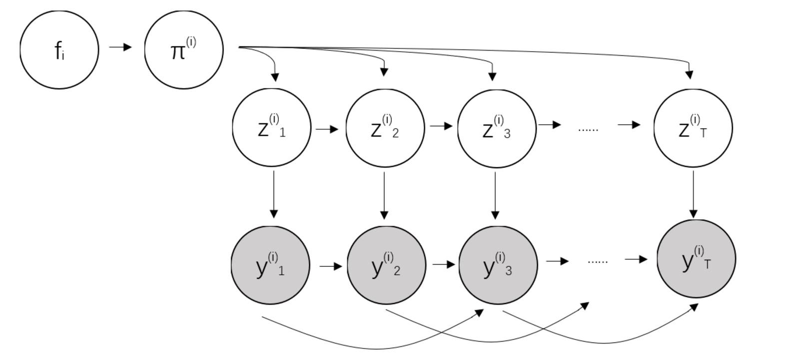

Figure 6: Graphical model of GAR HMM, Gaussian auto-regression hidden Markov.

By applying the auto-regressive hidden Markov model on 16 behaviors, the probability changes over time can be modelled. Figure 7 shows an example the probabilistic changes of eye gaze in one region.

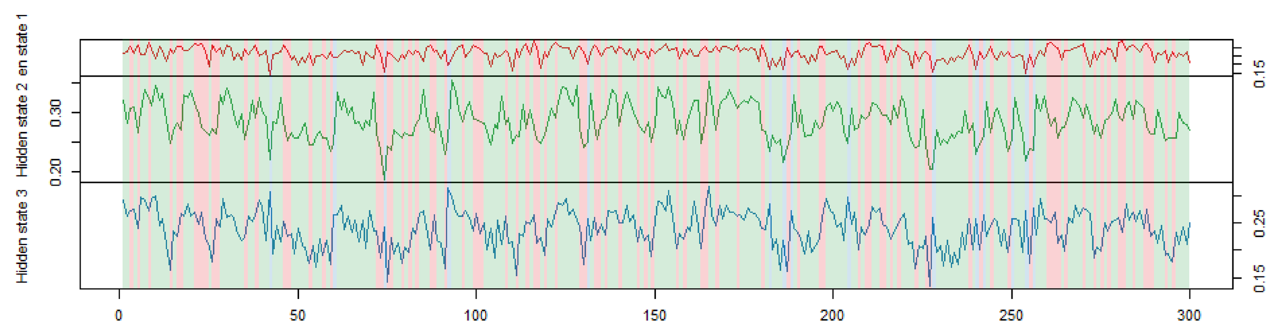

Figure 7: Example of behavior change and and the probability change presented a graph over time. Three hidden states are presented in three colours.

Result

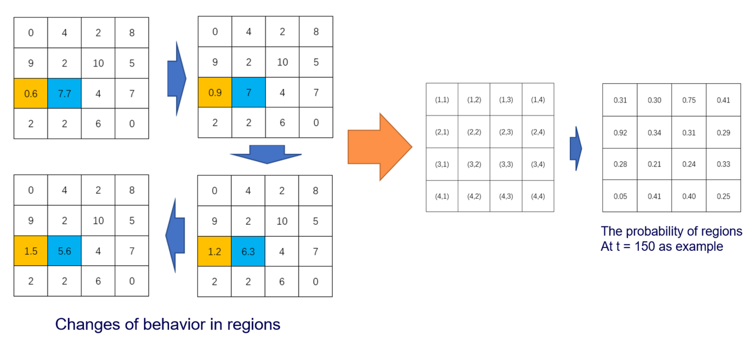

Figure 8: Transform behavior matrix into probability distribution.

Each region represents the original data of the potential behavior of the eye-gaze movement in the image region, and this behavior will share changes with other regions. By importing the 4×4 behavior information contained in the regions of the image into GAR HMM, we can get 16 probability change curves over time, which represent the potential possibility of eye gaze movement in 16 regions of the image, and generate probability distribution over time.

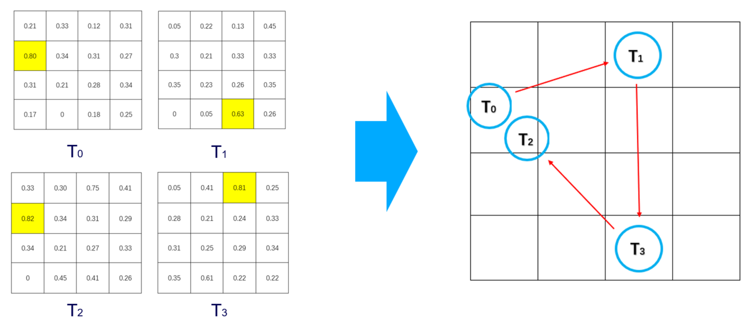

Figure 9: As an example, demonstration of predictive eye gaze movement over time ($t_0 : t_3$)

Using generated probability distributions, the instantaneous tendency of the human eye can be captured. The eye gaze will move toward the high probability area. By selecting the region with the highest probability for each time period, we get a time series of eye gaze movement predictions by the coordinates of regions.

Figure 10 represents the average test accuracy of prediction from different types of partition. When the number of split regions of the image increases, the accuracy of the probability prediction will increase as well, except 5. With a 4×4 grid GAR HMM can achieve good accuracy in capturing the eye movement tendency, which is 73.7%. If the partition is too coarse, for example 2×1, the accuracy is close to random guess, 50%. That meas two regions are insufficient to capture the characteristic of eye gaze movement. We hold the hypothesis that when the partitions are within a computationally feasible range, the more partitions, the more potential eye gaze movement information captured by the Gaussian auto-regression hidden Markov model, so the eye gaze movement prediction can be more accurate.

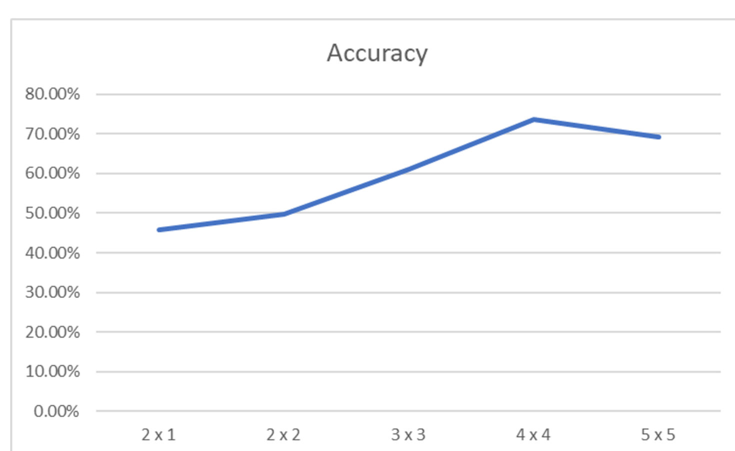

Figure 10: Average prediction test accuracy using different types of partition

Summary

In this research, we proposed a filtering method to remove non-principle movement by k-Nearest Neighbor classification and Fourier transform.The processed trajectories after the filtering are consistent with the underline image content. We also proposed a new method, Gaussian auto-regression hidden Markov model to predict how eye gaze moves between image regions base on the filtered eye gaze movement data. From our experiments, we show that the proposed predictive method (GAR HMM) is effective, in particular with region size of 4×4, where the test accuracy can be 73.7%.