Lost in Style: Gaze-driven Adaptive Aid for VR Navigation

Supplementary Material

ACM Reference Format:
. 2019. Lost in Style: Gaze-driven Adaptive Aid for VR Navigation: Supplementary Material. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3290605.3300578

1 MOTIVATION FOR USING GAZE ANGLES

We found that using gaze angles to be appropriate to predict need for navigation aid. After completing the data collection experiment, we compared using 3D and 2D gaze fixation sequences. We found that the average Dynamic Time Warping Euclidean distance between positive (windows with the navigation aid present) and negative examples (windows without a navigation aid present), was larger in 2D than 3D gaze fixation. The average normalized Dynamic Time Warping Euclidean distance was 62% and 39% for 2D and 3D gaze sequences respectively. These examples were created by randomly sampling 2000 windows from our training data with an equal number of positive and negative examples.

Three dimensional gaze points contain extraneous information (e.g. distance of the fixation from the screen) that are likely to mislead the predictor. Unlike the 3D gaze point which conveys information about the geometry of the scene, the 2D on-screen gaze point location is more informative of the user's gaze direction. Therefore, the RNN failed to predict the need for aid with adequate accuracy using 3D point sequences as well. The RNN was able to reach a maximum accuracy of 61% with the 3D gaze points.

In practice it is computationally more costly to compute the on-screen projected 2D gaze points than computing gaze angles. So we chose to use gaze points for the need for aid predictor.

CHI 2019, May 4–9, 2019, Glasgow, Scotland Uk

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300578

2 TOPOLOGY OF THE ADAPTIVE NAVIGATION AID CLASSIFIER

We did not use a deep network and thus we used a GPU for only training our recurrent neural network, but not for prediction. Figure 1 shows our network's topology which consisted of: an input layer of size 350 connected to a hidden layer consisting of 512 LSTM cells, an outer sigmoid activation layer with a binary cross entropy loss.

3 SPECIFYING THE AID REGION

Using a Wilcoxon Signed-rank test ($W = 18$, $p = 0.02$, $r = 0.62$), the region outside of the *permanent arrow* shown in Figure 2 received an average gaze fixation duration of (M=43.78, SD=16.21) seconds, significantly (27.04 seconds) shorter than the region outside of the *adaptive arrow* has received (M=70.82, SD=41.61).

To further investigate the effects of our adaptive arrow on gaze fixation duration patterns, we expanded the aid region to include some fixation points that could have been created due to participants viewing the aid using their peripheral vision. The expanded arrow region is shown in Figure 3. The area outside the *permanent arrow* rectangle received an average gaze fixation duration (M=31.9, SD=16.13) 17.75 seconds shorter than the area outside the *adaptive arrow* (M=49.65, SD=40.21) ($W = 20$, $p = 0.02$, $r = 0.59$). The Wilcoxon Signed-rank test indicated no significant difference in the average gaze fixation duration between our *adaptive min-map* and the *permanent mini-map*. The region outside of our *adaptive mini-map* shown in Figure 4 received an average gaze fixation duration of (M=52.4, SD=20.3) not significantly shorter than the average gaze fixation duration the *permanent min-map* received (M=70.37, SD=50.5) ($W = 51$, $p = 0.61$).

Expanding the mini-map region to include participant's peripheral vision did not change the Wilcoxon Signed-rank test's significance. The region outside of the expanded *adaptive mini-map* received an average gaze fixation duration of (M=34.26, SD=17.41) seconds, while the region outside the *permanent mini-map* received (M=50.3, SD=47.95) seconds ($W = 51$, $p = 0.61$).

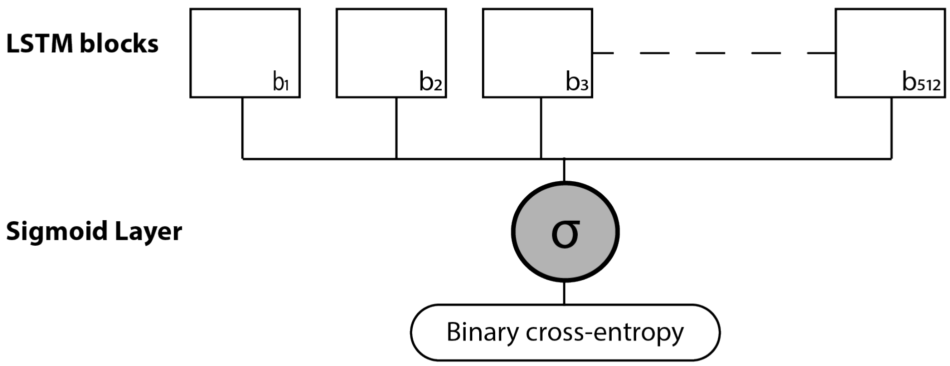

Figure 1: The topology of our adaptive navigation aid recurrent neural network. Which consists of 512 LSTM blocks with an input size of 350 (b_1, b_2, . . . , b_{512}), and a sigmoid activation layer with a binary cross-entropy loss.

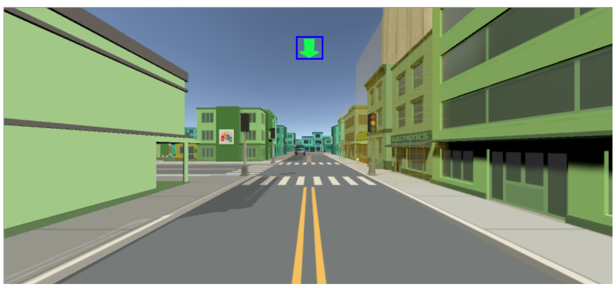

Figure 2: The blue rectangle shows the arrow region used for our outside the aid gaze fixation computation.

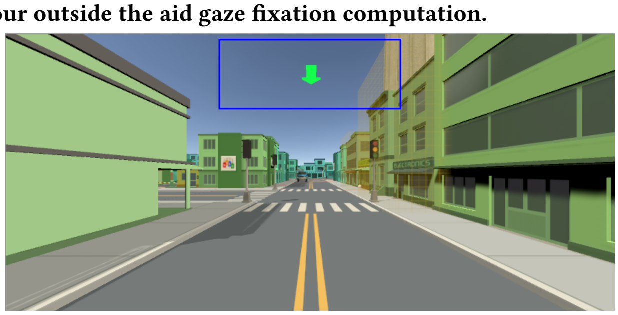

Figure 3: The blue rectangle shows the expanded arrow region used for our outside the aid gaze fixation computation. We expanded the rectangle to include peripheral vision gaze fixation points.

Figure 4: The blue rectangle shows the mini-map region used for our outside the aid gaze fixation computation.

Figure 5: The blue rectangle shows the expanded mini-map region used for our outside the aid gaze fixation computation. We expanded the rectangle to include peripheral vision gaze fixation points.