

ANALYSIS OF DRIVERS' HEAD AND EYE MOVEMENT CORRESPONDENCE: PREDICTING DRIVERS' GLANCE LOCATION USING HEAD ROTATION DATA

Mauricio Muñoz^{1,2}, Joonbum Lee¹, Bryan Reimer^{1*}, Bruce Mehler¹, and Trent Victor³
¹MIT AgeLab & New England Univ. Transportation Center, Cambridge, Massachusetts, USA
²TU Munich, LMU Munich, University of Augsburg, Munich, Germany
³SAFER Vehicle and Traffic Safety Centre at Chalmers, Gothenburg, Sweden.
*Corresponding author - reimer@mit.edu

Summary: The relationship between a driver's glance pattern and corresponding head rotation is not clearly defined. Head rotation and eye glance data drawn from a study conducted by the Virginia Tech Transportation Institute in support of methods development for the Strategic Highway Research Program (SHRP 2) naturalistic driving study were assessed. The data were utilized as input to classifiers that predicted glance allocation to the road and the center stack. A predictive accuracy of 83% was achieved with Hidden Markov Models. Results suggest that although there are individual differences in head-eye correspondence while driving, head-rotation data may be a useful predictor of glance location. Future work needs to investigate the correspondence across a wider range of individuals, traffic conditions, secondary tasks, and areas of interest.

INTRODUCTION

Eye movements have long been studied in the context of driver behavior, attention management, and to assess task related visual demand (e.g., Wierwille, 1993). Although eye tracking systems have been employed in numerous scientific studies (e.g., Wang, Reimer, Dobres & Mehler, 2014), the technology is susceptible to data quality issues (Ahlstrom & Victor, 2012) and is difficult to use in a reliable way in production systems. Research on the correspondence between eye and head movement suggests that head pose data may be useful as a surrogate for eye-glance data. Eye-head coordination has been investigated in several reports (e.g., Previc, 2000; Talamonti, Huang, Tijerina, & Kochhar, 2013). Previc (2000) found that the probability of a head turn accompanying an eye glance increases as the distance between the center of the road and a secondary point of focus increases beyond 20 degrees laterally. Similarly, Talamonti and colleagues (2013) found a low likelihood (65% or less) of head turns when glancing to the instrument panel and rearview mirror, and high likelihood (93% or more) when glancing to the left mirror, center console, and center stack. Talamonti (2014) suggests that driver-specific thresholds need to be set in order to meaningfully use head yaw data as a glance predictor. This study aims to further investigate whether head-rotation data can be used as a surrogate for certain eye-glance behaviors. Based upon the literature noted above, it was not expected that head-rotation data could be used to predict all glances away from the road. Therefore, as an initial effort, and approaching the problem from a classification perspective, we consider if head rotation can be used to predict glances to the forward road and to the vehicle's center stack (e.g., climate controls, infotainment display, etc.). Results are considered using both the original "skewed" dataset, which is characterized by a heavily uneven distribution of samples for each glance type (95% of all glances were forward glances), as well as a subset with an equal amount of glance samples for each type of glance. Further, in terms of model selection, six classifiers

that cover a wide range of data interpretation paradigms were examined. These steps allow us to reasonably begin to isolate the descriptive potential of the head pose signal.

METHODS

This study is a secondary analysis of data collected by the Virginia Tech Transportation Institute (VTTI) in support of methods development for the Strategic Highway Research Program (SHRP 2) naturalistic driving study (Transportation Research Board of the National Academies of Science, 2013). The data were provided to the MIT AgeLab under an IRB approved data sharing agreement. A total of 44 participants were included in the analysis (22 participants each for static and dynamic trials) spanning four age groups (18-35, 36-50, 51-65, 66+, with a majority falling in the first two groups) as well as a wide array of facial geometry. Approximately twice as many males (n=30) than females (n=14) were tested. Participants who met the study's eligibility criteria were assigned to participate in either static trials or dynamic trials. Data were collected in a 2001 Saab 9-3 instrumented with a data acquisition system (DAS) to collect a number of metrics including digital video of the drivers face. This video was recorded by a camera mounted below the rearview mirror.

Test Trials

Static testing occurred in a research lot at VTTI, during which participants were instructed to glance to predefined locations (e.g., "forward windshield", "center stack", etc.). The static testing consisted of nine prompted glance locations. Dynamic testing was conducted on a predefined route around Blacksburg, Virginia. During the session, participants were instructed to perform a set of five tasks (e.g., "report current vehicle speed", "turn the radio on and then off", etc.).

Data Reduction

For each video frame, geometric methods (e.g., Murphy-Chutorian & Trivedi, 2009) were used to generate pixel coordinates used for head rotation estimation. Each video frame was analyzed by two independent reviewers who annotated seven predefined facial landmarks. Glance locations were coded by trained video analysts into one of 16 locations (e.g., "forward", "instrument cluster", "center stack", etc.) on a frame-by-frame basis. A senior analyst reviewed the output of the coding and provided feedback. As this study only focuses on assessing the differences between standard forward glances and glances to the center stack, data for all other glance types were excluded. Glance allocations for each subject and task were merged with head rotation data using timestamps.

Model Training and Validation

Training data were derived from the dataset by taking all data belonging to a randomly sampled subset of all available subjects (80%). The remaining (20%) subjects were used to build a validation dataset. The timestamp ordering of the samples for each subject was maintained, as the Hidden Markov Models (HMMs) consider the temporal structure of the input data. In order to control for the effects of this random sampling method on classifier performance, a Monte-Carlo sampling technique was used. For each of 50 iterations, training and validation sets were generated, each of which were used to train all models. Performance values were then computed

for each classifier as the mean of each performance metric (accuracy, F1 score, and Kappa value) over all iterations. The raw dataset included more glances to the forward roadway (approx. 95% of glances) than to the center stack. Given that the skewness of a dataset (unbalanced class structure) influences classifier behavior, an alternative dataset was constructed where forward-glances were randomly removed until the number of glances to the two locations were equivalent.

Data Analysis

The intrinsic discriminative quality of the data plays a crucial role in any classification framework (classification may be difficult in datasets in which two classes overlap strongly in feature space). In order to explore this aspect of our data, visual representations of some of its most salient patterns were developed using principal component analysis (PCA). PCA is a common statistical technique for pattern detection, data visualization, and compression (Jolliffe, 2005). It is used here to represent the raw data in terms of its salient structural patterns (as given by the eigenvectors of its covariance matrix) and in order to visualize properties that might have an impact on classification performance.

MODEL DEVELOPMENT

Supervised classification is used in order to assess the predictive value of head rotation on glance allocations. Predictive models were sampled from a heterogeneous pool of approaches in order to examine which data interpretation method, if any, is best able to characterize the correspondence between head rotation and glance allocation data. In particular, models considered include feature-based distance metrics (k-Nearest Neighbor), clustering (Gaussian Mixture Models), boosting and ensemble methods (Boosted Decision Tree, Random Forests), monolithic classifiers (Neural Network), and statistical, temporal models (HMMs). There are various benefits and drawbacks to each modeling approach and a full description is beyond the scope of this work. In short, k-Nearest Neighbor predictors can be used to find spatial correspondence between samples. Experimental results show that looking at nine neighbors maximized classification performance. Gaussian Mixture Models (Multivariate) (Reynolds, 2009) are used to fit an additive combination of Gaussian distributions to the input variables, using one distribution to model each class. As an ensemble method, the two-class real AdaBoost (Friedman, 1998) algorithm (using 90 weak classifiers at a depth of 40) was used, as well as a related classifier, the Random Forest, with 100 trees in the forest, each with depth 50. A standard Multilayer Perceptron (Gardner, 1998) with three hidden layers was taken as a representative of the larger class of artificial neural networks for its ability to model non-linear relationships between data points. HMM (Rabiner, 1986) were used to analyze the discriminative potential of the temporal structure of the data. Empirical tests showed that using five hidden states maximized classification performance.

RESULTS

Group Level Analysis

The correspondence between head and eye movements is illustrated in Figure 1, where the distributions for individual mean horizontal head rotation (Y axis) are plotted independently for

forward glances and glances to the center stack across dynamic and static trials. The overlap between the two distributions during static trials is smaller than the overlap observed during dynamic trials.

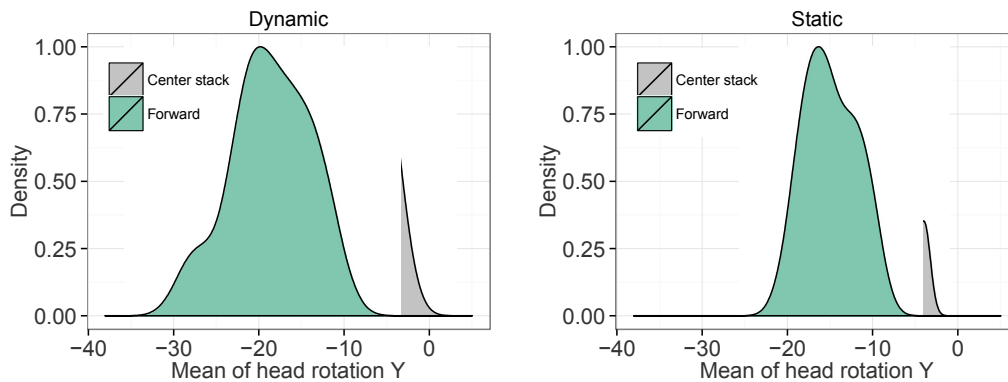


Figure 1. Distributions of head rotation X and Y while glancing forward and to the center stack

For the dynamic trials, the mean Y rotation for glancing to forward was -18.63 ($SD = 4.53$), and the mean Y rotation for glancing to the center stack was -8.56 ($SD = 3.83$). For static trials, the mean Y rotation for glancing to forward was -14.92 ($SD = 4.53$), and the mean Y rotation for glancing to the center stack was -6.34 ($SD = 1.43$). For the dynamic trials, the mean difference of head rotation X (vertical head rotation) while glancing forward and to the center stack was 6.25 ($SD = 3.64$) degrees, and the mean difference of head rotation Y (horizontal head rotation) was 7.72 ($SD = 3.47$) degrees. Individual differences in head-eye correspondence were also observed. For example, the minimum mean difference of head rotation X was 0.39 degrees, whereas the maximum was 12.35 . Likewise, the minimum mean difference of head rotation Y was 1.13 degrees, whereas the maximum was 14.09 degrees. The number of center stack examples in the static data summed up to less than half than the number of corresponding examples for the dynamic data. Preliminary experiments showed that the models were prone to over-fit the static data very easily, implying that not enough data were available for a fair comparison of model performance between the static and dynamic datasets. Therefore, the static data were not considered in the development of predictive models.

Principal Component Analysis

Figure 2 depicts the dynamic training data in terms of its principal components (PCA scores). Intuitively, each axis corresponds to a principal component and represents salient statistical behavior of the data along that component. A rough correlation between sample location and class could be established, implying there is at least one eigenvector/statistical pattern, which can potentially discriminate between classes (Figure 2a). The fact that there is no clear cut clustering suggests, however, that the problem at hand is not straight forward and that it is unlikely that any given classifier will perform extremely well. Taking a closer look at the first principal component (Figure 2b, first column) also reveals that both X and Y rotation features contribute significantly to the statistical variance (loosely, information content) of the underlying data. X rotation (up-down head movement) potentially qualifies as the most significant variable in the classification framework.

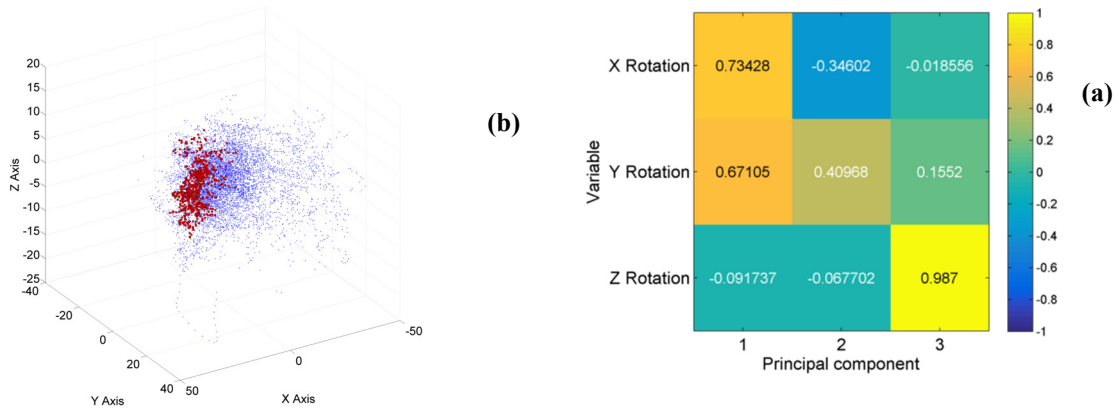


Figure 2. Principal component analysis (PCA) of dynamic data, using head rotation: (a) Glances to the center stack (red points) and forward roadway (blue points), (b) Principle components of head rotation x, y, and z

Model Validation

Experiments were conducted to assess the model performance across two principal factors, classifier type and skewness of the data (i.e., raw/balanced glance allocations to the two regions of interest). Three key performance measures are reported:

1. Classification accuracy (AC) (Sokolova & Lapalme, 2009): percentage of correctly classified samples (or sample sequences for the HMM classifier).
2. F1-score (FS) (Sokolova & Lapalme, 2009): a measure of how well the classifier was able to distinguish between classes given an unbalanced dataset.
3. Cohen’s Kappa statistic (KP) (Carletta, 1996): measures how well a classifier agrees with a perfect predictor (higher values indicate high agreement).

Table 1 presents performance measures for all six classifiers using the two representations of the data, computed using 50 iterations of the aforementioned Monte-Carlo sampling scheme. Using the balanced dataset that removes glance distribution bias during training leads to a higher performance in terms of sensitivity/specificity (F1 scores all ≥ 0.68) and prediction quality (Kappa all ≥ 0.41) of each classifier compared to (F1 scores all ≥ 0.04) and (Kappa all ≥ -0.07) for the original unbalanced data. The HMM classifies sample sequences corresponding to blocks of data within a subject and task group. The relatively strong performance of the model suggests that the temporal structure of head rotation features is another potential source of information. Overall, there is a general consensus amongst all classifiers regarding the discriminative quality of head rotation data. Though all classifiers using the balanced dataset perform better than a chance predictor, there is a clear upper bound on how much these features contribute to classification. Though PCA reveals similar variance in both X and Y rotations, the Random Forest classifier found Y rotation to be most relevant variable to the classification efforts.

Table 1. Performance measures (accuracy, F1 score, Kappa statistic) across all classifiers and class distributions for dynamic data

	Original Dataset			Balanced Dataset		
	AC	FS	KP	AC	FS	KP
k-Nearest Neighbor	0.93	0.19	0.16	0.80	0.80	0.59
Gaussian Mixture Model	0.56	0.04	-0.07	0.73	0.74	0.46

Boosted Trees	0.90	0.19	0.14	0.70	0.68	0.41
Random Forest	0.86	0.30	0.24	0.79	0.78	0.59
Multilayer Perceptron	0.78	0.29	0.22	0.80	0.82	0.59
Hidden Markov Model	0.84	0.28	0.22	0.83	0.68	0.57

DISCUSSION AND CONCLUSION

This study investigates the relationship between head rotation and glance behavior during static and on-road driving. The distributions of head rotations (Figure 1) illustrates divergent relationships between head position and eye glance locations while driving compared to behaviors in a stationary car. In particular, using only glances to the center stack and the forward roadway, the dynamic trials showed more overlapping of head rotation range in comparison to the static trials. This implies that drivers were more likely to keep their heads oriented towards the road while glancing to the center stack, and that movement data generated in non-driving applications may not be useful in the development of models aimed to predict eye movements. Machine-learning techniques were employed to examine the predictive value of head rotation for glance location at an individual level. A total of six classifiers from a wide range of data interpretation techniques were used to detect patterns in head rotation data. Substantial performance gains were observed when using a balanced training dataset. HMMs performed the best with accuracy of 83%. All of the modeling approaches provided results that were well in excess of chance findings, suggesting that head rotation data is a fairly robust predictive signal. Given the limited number of glances to the center stack (<5% of the total glances recorded) captured during short, simple tasks, model performance may be best considered as a lower bound on the possible predictive quality. Given the time series nature of more complex glance allocation strategies and performance of the HMM, higher predictive accuracies may be achievable. The scope of this study was limited to classifying only two glance objects (i.e., forward and center stack) in a previously collected and annotated dataset. Subsequent work may wish to consider model performance across a larger and perhaps more spatially diverse set of glance objects with data drawn from a larger population of drivers. Further, efforts should assess the predictive power of head rotation data for certain types of glances such as those that are of longer duration and linked to greater risk of collision (Victor, 2014). Overall, this work suggests that head rotation data, a feature that can be recorded with commercially available sensors, may provide a potentially lower cost and reasonable estimate of attention allocation compared to eye tracking data. Further study of head movements as a means to predict safety critical off-road glances to regions frequently associated with in-vehicle distractions appears warranted.

ACKNOWLEDGEMENTS

Support for this work was provided by the US DOT's Region I New England University Transportation Center at MIT and The Santos Family Foundation.

REFERENCES

Ahlstrom, C., Victor, T., Wege, C., & Steinmetz, E. (2012). Processing of eye/head-tracking data in large-scale naturalistic driving data sets. *Intelligent Transportation Systems, IEEE Transactions on*, 13(2), 553-564.

- Carletta, J. (1996). Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, 22(2), 249-254.
- Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The Annals of Statistics*, 28(2), 337-407.
- Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14-15), 2627-2636.
- Jolliffe, I. (2005). *Principal component analysis*. John Wiley & Sons, Ltd.
- Murphy-Chutorian, E., & Trivedi, M. M. (2009). Head pose estimation in computer vision: A survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(4), 607-626.
- Previc, F. (2000). Neuropsychological guidelines for aircraft control stations. *Engineering in Medicine and Biology Magazine, IEEE*, 19(2), 81-88.
- Rabiner, L., & Juang, B. H. (1986). An introduction to hidden Markov models. *ASSP Magazine, IEEE*, 3(1), 4-16.
- Reynolds, D. (2009). Gaussian mixture models. *Encyclopedia of Biometrics*, 659-663.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427-437.
- Talamonti, W. J., Huang, W., Tijerina, L., & Kochhar, D. (2013). Eye glance and head turn correspondence during secondary task performance in simulator driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 57(1), 1968-1972.
- Talamonti, W. J., Kochhar, D., & Tijerina, L. (2014). Eye glance and head turn correspondence during secondary task performance in simulator driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58, 2224 -2228
- Transportation Research Board of the National Academies of Science. (2013). *The 2nd strategic highway research program naturalistic driving study dataset*.
- Victor, T., Bårgman, J., Boda, C. N., Dozza, M., Engström, J., Flannagan, C., ... & Markkula, G. (2014). *Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk* (No. SHRP 2 Safety Project S08A)
- Wang, Y., Reimer, B., Dobres, J., & Mehler, B. (2014). The sensitivity of different methodologies for characterizing drivers' gaze concentration under increased cognitive demand. *Transportation Research Part F: Traffic Psychology and Behaviour*, 26, 227-237.
- Wierwille, W. W. (1993). Visual and manual demands of in-car controls and displays. In B. Peacock & W. Karwowski (Eds.), *Automotive Ergonomics* (pp. 299-320). London: Taylor and Francis.