Head and Gaze Dynamics in Visual Attention and Context Learning

Anup Doshi and Mohan M. Trivedi
Laboratory for Intelligent and Safe Automobiles
Computer Vision and Robotics Research Lab
University of California, San Diego
{andoshi,mtrivedi}@ucsd.edu.

Abstract

Future intelligent environments and systems may need to interact with humans while simultaneously analyzing events and critical situations. Assistive living, advanced driver assistance systems, and intelligent command-and-control centers are just a few of these cases where human interactions play a critical role in situation analysis. In particular, the behavior or body language of the human subject may be a strong indicator of the context of the situation. In this paper we demonstrate how the interaction of a human observer's head pose and eye gaze behaviors can provide significant insight into the context of the event. Such semantic data derived from human behaviors can be used to help interpret and recognize an ongoing event. We present examples from driving and intelligent meeting rooms to support these conclusions, and demonstrate how to use these techniques to improve contextual learning.

1. Introduction

Those Intelligent Environments that must assist humans in time- and safety-critical situations would benefit from using contextual information in recognizing objects and events. Such support systems could observe multimodal cues from humans as well as the environment to robustly and efficiently analyze the scene. Context information that could be derived from human behavior include, for example, whether there are distractions, or what kinds of goals the humans have in mind. Such cues might manifest themselves in various human behaviors under different contexts. This leads to two related questions:

1. Is it possible to observe differences in human behavioral cues in different contexts?
2. Is it possible to extract and use those cues to help identify or learn contextual information?

In particular, the task of driving offers many avenues for intelligent systems to assist people in improving safety and

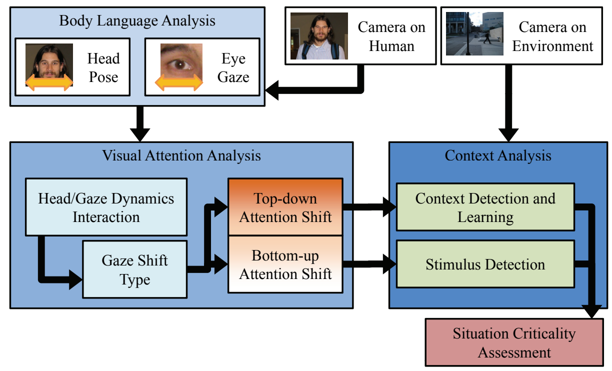


Figure 1. Flowchart of proposed approach for context and situation analysis. The body language analysis is used to derive attentional behavior, which can give important clues to the context analysis engine.

performance. Every year there are over one million traffic-related fatalities worldwide [27], with an estimated 26,000 in the U.S. alone due to driver inattention [16, 1]. Recent advances have promoted the integration of driver behavior analysis into Intelligent Driver Assistance Systems [36, 35], to counteract inattention and poor driving behavior. The analysis of body language in critical situations, such as the time prior to a lane change, becomes central in giving these systems the ability to predict the context of the situation.

Another environment in which an intelligent system could benefit from human behavior-related context analysis is in intelligent Command-and-Control centers or intelligent meeting rooms. An interactive system monitoring the participants or operators, could base decisions on context and amount of required assistance upon the subjects' body language. This may help reduce distractions and help improve performance of whatever task is being performed.

For example, if an operator is checking a particular monitor for some expected information, they may have different body language than when they are distracted by something or if something on the monitor draws their attention unex-

978-1-4244-3993-5/09/$25.00 ©2009 IEEE

pectedly. Landry *et al*. [15] discovered that certain patterns, or "gestalts", of aircraft on a radar screen drew the attention of the air traffic controllers due to their location, though they were not relevant to the task. An air traffic control training manual from the FAA [4] states that "even in low-workload conditions, distractions can clobber short-term or working memory." An assistance system could mitigate this dangerous situation with knowledge of the context of the attention shift, with warnings when the attention shift is not task-related.

In this paper we demonstrate how the interaction of an observer's head pose and eye gaze behaviors can provide significant insight into the context of the event, as seen in Figure 1. In essence, we claim that one can detect whether the human subject had prior knowledge of the scene or the task upon which they are focusing. That contextual knowledge may be useful to real-time interactive object and event recognition systems. Specifically, it could provide semantic hints such as whether there is a distraction in a scene, or what kinds of goals the human subjects are pursuing. We present examples from driving and intelligent meeting rooms to support these conclusions.

The remainder of the paper is organized as follows. Section 2 covers previous work, and an overview of gaze behaviors is presented in Section 3. Supporting analysis from an intelligent meeting room-style environment is in Section 4, with further analysis from a driving-related environment in Section 5. A concluding discussion appears in Section 6.

2. Related Research

2.1. Human Behavior Analysis and Prediction

A great amount of recent research has been in the detection of human behaviors. Many of these examples use patterns of human behavior to learn about the scene or predict future behaviors [33, 26, 29]. In considering time- and safety-critical situations such as driving scenarios, we present an overview of research related to behavior analysis and prediction in vehicles.

2.1.1 Driver behavior and Intent Prediction

Recent research has incorporated sensors looking inside the vehicle to observe driver behavior and infer intent [36, 35]. Bayesian learning has been used to interpret various cues and predict maneuvers including lane changes, intersection turns, and brake assistance systems [6, 21, 20]. More recently head motion has been shown to be a more useful cue than eye gaze for discerning lane change intentions [7].

The assumption made in all of these systems has been that head motion or eye gaze is a proxy for visual attention. In other words the system tries to measure head motion given that the driver is likely paying attention to whatever they are looking at, in whichever direction they are looking. The system then infers that because their attention is in a certain direction, they must have goals associated with that direction. For example, a driver may look left prior to changing lanes, as a direct result of their need to be attentive of vehicles in the adjacent lane.

2.2. Gaze behavior and Visual Search

However a gaze shift may or may not be associated with a particular goal. The broad question of "why people look where they look" is a subject of research for cognitive psychologists, neuroscientists, and computer vision researchers alike.

A significant amount of research in psychology has examined whether such visual searches are guided by goals (such as the goal of changing lanes), or by external stimuli [12, 25, 30, 13, 32]. These stimuli may include visual distractions, which could pop up in a scene and thereby attract the attention of the observer. For the most part any visual search is presumed to be guided by some combination of a goal-driven and stimulus-driven approach, depending on the situation [39].

Itti *et al*. [12, 25, 30] have made inroads in developing saliency maps that model and predict focus of attention on a visual scene. Initial models were based on "bottom-up" based cues, such as edge and color features. However it was found that "top-down" cues may be more influential; in other words the *context* of the scene and the goals of the human subject are crucial to determining where they look. Jovanevic *et al*. [13] determined that even in the presence of potentially dangerous distractions, in complex environments gaze is tightly coupled with the task. Several other works have similarly concluded that in natural environments, saliency does not account for gaze, but task and context determine gaze behavior [14, 32, 11, 41].

It is clear that in certain critical environments, distractions play an important role in attracting attention. Carmi and Itti [5] show that dynamic visual cues play a causal role in attracting visual attention. In fact, perceptual decisions after a visual search are driven not only by visual information at the point of eye fixation but also by attended information in the visual periphery [10]. In certain cases these stimuli may affect task performance; Landry *et al*. [15] found certain unrelated gestalt motion patterns on radar screens drew the attention of air traffic controllers away from the task at hand. In the driving context there are many well-known cognitive and visual distractions that can draw the driver's attention [1]. Recarte and Nunes [31] measured the number of glances to the mirror during a lane change, noting that visual distractions decrease the glance durations from by 70-85%. This result is well-aligned with more recent results indicating the limitations of drivers' multi-tasking abilities [18]. Moreover, some suggest that visual distractions may even increase likelihood of "change blindness", a phenomena whereby a subject may look in a certain area and not see or comprehend the objects in front of them [9, 17]. In these cases, it would be useful to know whether a gaze shift is attributable more to irrelevant visual stimuli or to a specific goal or context.

Several studies have used eye gaze or head pose to detect the attention of the subject [3], or estimate user state or gestures [2, 19]. As shown in Figure 1, we instead use the interaction of eye gaze and head pose to determine the attentional state of the subject, and proceed to use that information as contextual input to event detection and criticality assessment systems. In the following section we describe the framework of eye-gaze interactivity analysis, and in the later sections we describe some supporting experiments.

3. Framework for Proposed Approach

3.1. Attention shifts: top-down versus bottom-up

Attention shifts can be of two kinds, or some combination of the two [39]. Top-down attention shifts occur when the observer has a particular task in mind. This task may necessitate a visual search of a potentially predetermined location, or a search of likely relevant locations. An example may be as simple as shifting one's attention from a television to a newspaper, after having turned the television off. There may also be learned tasks, such as the search for oncoming cars when crossing a road, as a driver in a vehicle or as a pedestrian at a crosswalk.

Bottom-up attention shifts are caused by interesting stimuli in the environment. These stimuli may include distractions or salient regions of a scene. For example, flashing police lights on the highway may draw unnecessary attention, as may an instant chat message popping up on the screen during a technical presentation.

In many cases an object in the scene may easily be a distraction in one instance, and part of the task at hand at another time. For example, in a classroom, the person standing in front of the blackboard may be the teacher during a lesson, to whom the student should be paying attention. On the other hand, the person in front of the blackboard may just be someone walking by, who is distracting the student from the task at hand. By classifying the student's interactive behaviors, we may be able to easier set the context for each object and thereby improve recognition.

3.2. Head and Eye Gaze during attention shifts

Zangemeister and Stark [40] performed a controlled study of eye-head interactions and posited various conditions of different styles of eye-head movements. In their paper they found several styles of movements, depicted in Figure 3.2. Among the most pertinent of movements are

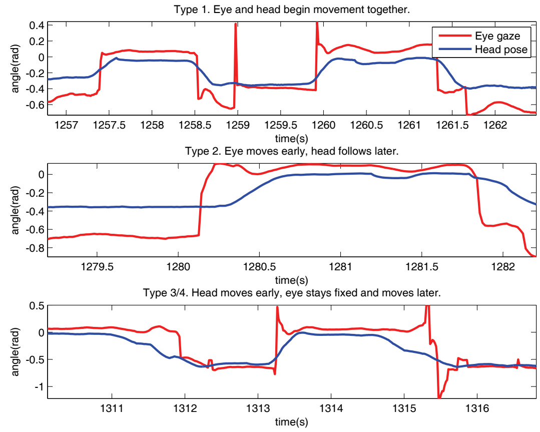


Figure 2. Examples of various interactions of head and eye movements, with type labels from [40]. Note in certain cases eye gaze tends to move first, where in others the head tends to move first.

those labeled "Type III," which include early or anticipatory head movement with respect to the gaze shift. They theorized that this behavior is associated with a repetitive, predetermined, or premeditated attentional shift, as is the case for any goal-directed attentional shift.

Morasso *et al.* [22] examined control strategies in the eye-head system, and observed that "The head response [to a visual target], which for random stimuli lags slightly behind the eyes, anticipates instead for periodical movements [of the target]." The implication is once again that for trained or predetermined gaze shifts, the head movement anticipates the gaze.

Indeed eye shifts can occur much faster than head movements and thus near-instantaneous gaze changes can be made with eye shifts alone. Shifts of larger amplitudes would necessitate a head movement to accommodate the limited field of view of the eyes. Most likely, when a large visual attentional shift is about to occur, and the observer has prior knowledge of the impending shift, these studies imply that there may be some amount of preparatory head motion [28].

In the following sections, we show that by extracting the dynamics of eye gaze and head pose, it may be possible to identify those gaze shifts which are associated with premeditated or task-oriented attentional shifts. In each case, we find that a majority of task-related shifts occur with an anticipatory Type III gaze shift. Based on these results and the studies listed above, we might further hypothesize that a Type III gaze shift could imply a task-related shift, and Types I or II are more likely to occur in conjunction with stimulus-related and smaller gaze shifts.

4. Experiments with Visual Displays

In the case of intelligent rooms such as Command-and-Control centers or Meeting rooms, it may be useful to analyze the behavior of the participants in order to sense the context of the interaction. For example, if the focus of attention shifts to a projection screen at some point, it may be due to a direction by some participant, or due to a distraction. In such a case, if we are able to detect a Type III visual search by a participant, it may indicate that the attention shift was goal-oriented and thus the material on the screen may be relevant to the current discussion. The context may be different if the participant is bored or distracted, in which case an ambient sensor may be directed to find the source of the distraction.

In this section we present the results of a meeting simulation study in which we analyze the gaze behavior of the participants in various situations. The results will demonstrate the feasibility of detecting body language to help infer events and the current context of the meeting.

4.1. Data collection

The setup in this experiment is designed to simulate a tele-conference or meeting presentation, where there is one main speaker with slides as additional material. This could of course be repeated with real people as opposed to videos of people, as in a meeting room scenario. However to maintain repeatability, videos are used in these trials.

As shown in Figure 4.1, the subject is seated facing a main monitor displaying a video of a presenter; and their slides show up on a side monitor sitting on the left side of the participant. The video is simply one of a professor lecturing, with the corresponding presentation slides shown on the side monitor.

Participants are instructed to pay attention to the video presentation as they might do naturally. They are also directed, as a secondary task, to glance over at the slides on the side monitor whenever the video professor indicates to do so (by pointing or looking over his shoulder). This gaze-following/side-monitor-checking task would simulate a goal-oriented attention shift in the subject.

At various other times during the experiment, an attempt is made to distract the subject from their task, without the subject's knowledge. These distractions include flickering small LED's, abruptly changing the display on the secondary monitor, and arbitrarily moving into the subject's peripheral field of view to halt the experiment. These distractions are chosen simply because they are capable of drawing the subject's visual attention away from the task at hand. Whenever the distraction is successful at engaging the visual attention of the subject, a note is made of that time.

Figure 3. Setup of the Intelligent Room experiment. Note the subject has to maintain focus on the primary monitor while checking the secondary monitor under certain conditions. The head and eye motion is tracked using a commercial eye tracking system.

It is important to note that a thorough treatment of distractions would require controls for numerous variables including the awareness level of the subject, which is beyond the scope of these preliminary experiments. Future work would include measuring and accounting for such variables to gain further insight. However currently we find that it is enough to note the existence of a reaction to a visual distraction; when there is no reaction we simply ignore that example in analysis. As seen in the results in Table 4.2, this artificially deflates the number of False Negative detections, although we are able to draw interesting conclusions even without those results.

During the entire experiment, eye gaze and head pose are captured by using a commercially-available non-intrusive stereo-camera eye tracking system. The system uses infrared illumination and requires manual calibration for each

Table 1. Confusion Matrix for Detecting Goal-Oriented Gaze Shifts (G) versus Stimulus-Oriented Shifts (S) in Meeting-Room Experiment

Number of examples...		
	Actually G	*Actually S*
Predicted G	46	0
Predicted S	6	9

subject. In addition to outputting data related to eye and head dynamics, it also outputs a confidence gauge in its own estimates. To ensure reliability results are retained only when this measure remains above a significant threshold.

4.2. Analysis

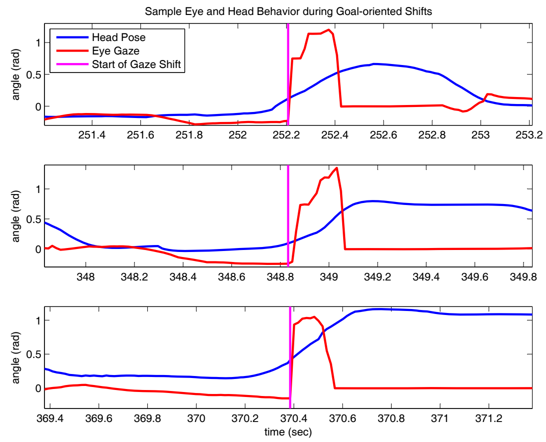
A total of six subjects were tested in this experimental setup, each on 10 to 15 minutes of lecture-style video. Out of all the data, 61 usable gaze shifts were found, as others were discarded due to low tracking confidence or noise. These gaze shifts were then manually labeled as either 'stimulus-oriented', if there was a distraction was present during the gaze shift, or 'goal-oriented'.

According to the hypothesis presented above, there may be a particular interaction between eye gaze and head pose, indicative of a goal-oriented gaze shift. Specifically, in the Type III case, a preparatory head motion may occur before the actual gaze shift. In order to detect such a motion, if it exists, we first must find the exact locations of the start of the actual gaze shifts. These shifts occur at an inflection point in the eye gaze signal, which is found as the maximum of the second derivative of eye gaze signal. To ensure robustness we only search for second derivative maxima in a small window around peaks in the raw eye gaze signal.

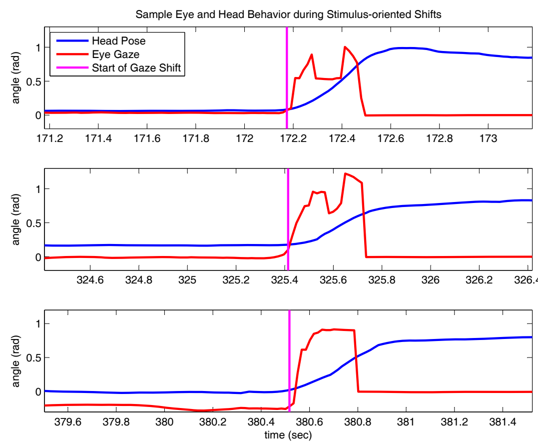
We then measure the average head motion, over a .5-second window prior to the initial gaze shift. If this motion is relatively small, then we might conclude there is no preparatory movement; on the other hand a large motion vector might likely indicate a Type III interaction associated with goal-oriented gaze shifts.

Optimal results were obtained by setting a threshold of approximately .15 radians per second. As shown in Table 4.2, 46 goal-oriented gaze shifts were correctly identified, for a detection rate of 88.5%. The stimulus-oriented shifts were all correctly classified as well, though in smaller numbers.

A closer look at some of the raw data gives some interesting insight. The top of Figure 4 shows several examples of the eye gaze and head pose data from goal-oriented gaze shifts, while the bottom shows some stimulus-oriented shifts, from various subjects in this meeting-style experiment. An intelligent environment may be able to spot the body language of these subjects and either assist them in their task, or gain clues on how to detect and reduce distrac-



(a) Samples of eye-head behavior during goal-oriented shifts



(b) Samples of eye-head behavior during stimulus-oriented shifts

Figure 4. Various examples of goal-oriented (a) and stimulus-oriented (b) gaze shifts obtained from an interactive meeting room-style experiment. Note the preparatory head motion prior to the goal-based shifts.

tions.

5. Experiments with Naturalistic Driving

McCall *et al.* [21] demonstrated the ability to detect a driver's intent to change lanes up to 3 seconds ahead of time, by analyzing driver head motion patterns. Doshi and Trivedi [7] extended this study and found that head motion was in fact an *earlier* predictor of lane change intentions than eye gaze. However the reasons for this interesting find were not clear.

We propose that the visual search that occurs prior to lane changes, and potentially in other similar common driving maneuvers, is initiated by a top-down process in the driver's mind. The driver has a goal in mind, and thus is trained to initiate a search in particular locations such as the mirrors and over the shoulders, for obstacles. Here we present a deeper analysis into real driving data to support this hypothesis, that the visual search prior to lane changes is a Type III search. The ability to detect this type of behavior is crucial in being able to identify the context of the situation, and then to assess its criticality or determine if objects around the vehicle are of interest.

5.1. Data collection

For this research, data was collected in a driving experiment with an intelligent vehicle testbed outfitted with a number of sensors detecting the environment, vehicle dynamics, and driver behavior. This data is drawn from the same data as was used in the lane change intent work by McCall *et al*. [21]. A camera-based lane position detector and CAN-Bus interface provided most of the data related to the vehicle and surrounding environment.

The main driver-focused sensor was a rectilinear color camera mounted above the center console facing toward the driver, providing 30fps at 640x480 resolution. To calculate head motion, optical flow vectors were compiled and averaged in several windows over the driver's face (detected with a Viola-Jones [37] face detector). This method was found to be stable and robust across different harsh driving conditions and various drivers. Other methods could be used for these purposes [24, 23]. Various automatic eye gaze detectors exist (e.g., [38]), however to ensure accuracy and reliability, eye gaze was labeled using a manual reduction technique similar to several recent NHTSA studies on workload and lane changes [16, 1].

The dataset was collected from a naturalistic ethnographic driving experiment in which the subjects were not told that the objective was related to lane change situations. Eight drivers of varying age, sex, and experience drove for several hours each on a predetermined route. A total of 151 lane changes were found on highway situations with minimal traffic. 753 negative samples were collected, corresponding to highway "lane keeping" situations.

5.2. Analysis

These examples were used with the features described above to train a classifier to predict 3 seconds in advance of a lane change, whether the driver would intend to change lanes. Another classifier was train for 2 seconds ahead of the lane change. In these studies an extension of SVM, namely Relevance Vector Machines, was used to classify to enforce sparsity and account for fewer training examples [34, 21]. In a comparative study, Doshi *et al*. [8] found that such a classifier based on head motion has significantly more predictive power than one based on eye gaze 3 seconds ahead of the lane change, but not 2 seconds ahead of time.

Table 2. Average Intent Prediction Confidences ($\overline{IPC}$) for Each Type of Classifier

Seconds before Lane Change:	*3 Sec*	*2 Sec*
Eye-Gaze Classifier ($\overline{IPC_{eye}}$)	0.27%	46.91%
Head-Pose Classifier ($\overline{IPC_{head}}$)	44.11%	66.39%
ANOVA: $\overline{IPC_{head}} > \overline{IPC_{eye}}$	$p < .01$	$p > .05$

By looking at the outputs of each classifier, we can get a better sense of the performance of eye gaze and head pose over time. Specifically, the RVM classifier outputs a class membership probability, ranging from -1 (for negative examples) to 1 (for positive examples); thus the more positive the value, the more confident it is in its predictions of a true intention. By looking at the average over all positive examples of these "Intent Prediction Confidences," in Table 5.2, we can tell that the Eye-Gaze-based classifier is hardly better than chance 3 seconds before the lane change, but improves significantly in the 2 second case. On the other hand, the Head-Motion-based classifier works very well even in the 3 second case.

The results suggest that drivers engage in an earlier preparatory head motion, before shifting their gaze to check mirrors or blind spot. A number of samples of eye gaze vs. head pose can be seen in Figure 5, with the top of that figure showing actual pictures of the driver at various points through the maneuver. The preparatory motion can certainly be seen in many of these cases.

ANOVA significance tests comparing the population of Intent Prediction Confidences ($\overline{IPC}$) demonstrate quantitatively that the preparatory head motion prior to the eye gaze shift, is a significant trend across all drivers ($p < 0.01$). This implies that the detection of a Type III gaze shift may be quite useful in future Advanced Driver Assistance Systems, by helping to determine the context of the drive and whether the driver is indeed paying attention or not.

6. Concluding Remarks

There are many arenas for intelligent systems to significantly improve safety and comfort of everyday life. Intelligent systems and agents can monitor the environment and gain contextual awareness of a situation in order to assist people in performing a task, or even to perform the task independently. In many situations, the interactions of human subjects is an essential key to understanding the environment and the context of the situation. Real-time object or event detection systems could find semantic data derived from human behaviors useful when interpreting a visual scene.

A particular aim of this project has been to use computer vision to determine whether a visual search is more influenced by a top-down goal or a bottom-up stimulus. Existing research hypothesizes, as the success of previous intent de-

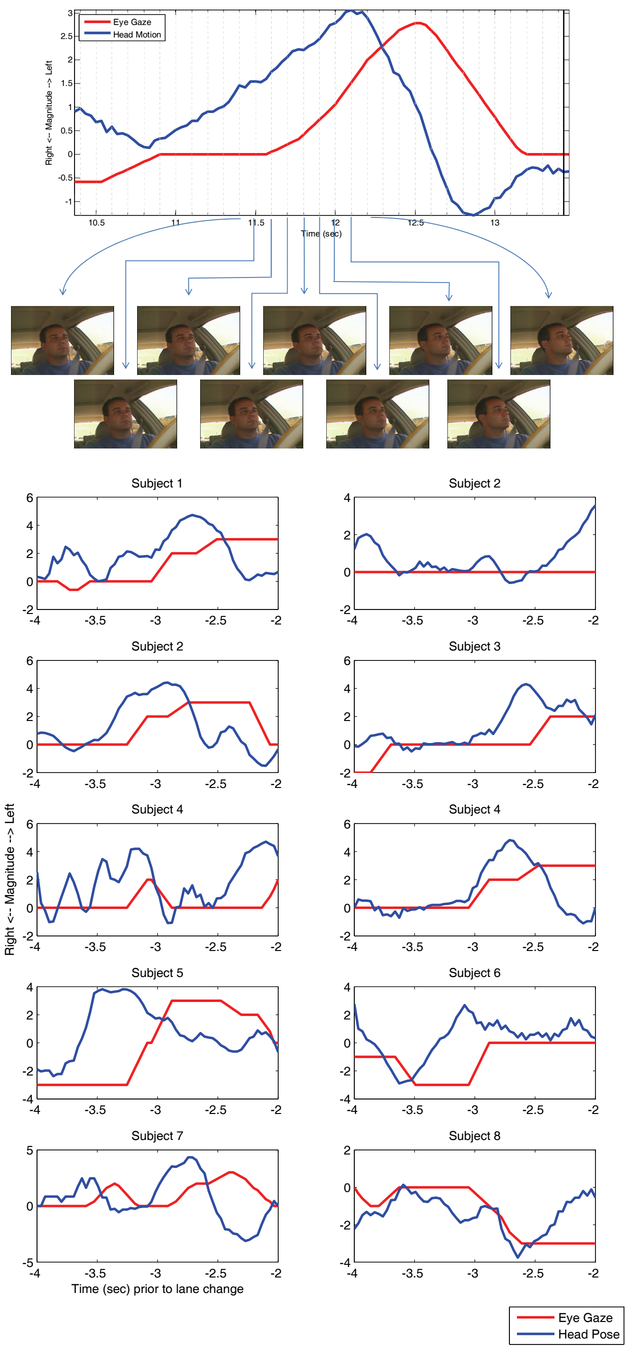


Figure 5. Various samples of eye-head motion prior to lane changes in several test subjects. Images from an actual naturalistic driving experiment are shown above, with a depiction of each stage of the gaze shift where the head moves first.

tection models reasonably demonstrate, that while performing a specific task, most gaze changes will be associated with the task-oriented visual search [32]. However distractions are a well documented and growing risk in driving and other environments, and so external stimuli may be increasingly competitive with driving tasks in attracting attention. Intelligent environments would benefit from the knowledge of the cognitive processes behind attentional shifts in humans, as such contextual clues could direct the system to assist or act appropriately in critical situations.

We have quantitatively demonstrated the ability to distinguish goal-oriented behaviors in humans by analyzing the interaction of head and eye movements. These results are seen in experiments from an intelligent room and intelligent vehicle scenarios. By distinguishing such body language we are better able to draw out the context of the scene, and whether that context is time- or safety-critical. Intelligent environments could use this context to assist humans and improve the safety and comfort of their interactions and surroundings.

References

[1] L. Angell, J. Auflick, P. A. Austria, D. Kochhar, L. Tijerina, W. Biever, T. Diptiman, J. Hogsett, and S. Kiger. Driver workload metrics task 2 final report. *Report DOT HS 810635, NHTSA, U.S. Department of Transportation*, Nov 2006.

[2] S. Asteriadis, P. Tzouveli, K. Karpouzis, and S. Kollias. Estimation of behavioral user state based on eye gaze and head poseapplication in an e-learning environment. *Multimedia Tools and Applications*, 41(3):469–493, February 2009.

[3] J. P. Batista. A real-time driver visual attention monitoring system. *Iberian Conference on Pattern Recognition and Image Analysis*, 2005.

[4] K. M. Cardosi. Human factors for air traffic control specialists: A user's manual for your brain. *Report DOT/FAA/AR-99/39, Federal Aviation Administration, U.S. Department of Transportation*, 1999.

[5] R. Carmi and L. Itti. Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research*, 46(26):4333–4345, December 2006.

[6] S. Y. Cheng and M. M. Trivedi. Turn intent analysis using body-pose for intelligent driver assistance. *IEEE Pervasive Computing, Special Issue on Intelligent Transportation Systems*, 5(4):28–37, October-December 2006.

[7] A. Doshi and M. M. Trivedi. A comparative exploration of eye gaze and head motion cues for lane change intent prediction. *IEEE Intelligent Vehicles Symposium*, June 2008.

[8] A. Doshi and M. M. Trivedi. Investigating the relationships between gaze patterns, dynamic vehicle surround analysis, and driver intentions. *IEEE Intelligent Vehicles Symposium*, June 2009.

[9] P. J. Durlach. Change blindness and its implications for complex monitoring and control systems design and operator training. *Human-Computer Interaction*, 19:423–451, 2004.

[10] M. P. Eckstein, A. Caspi, B. R. Beutter, and B. T. Pham. The decoupling of attention and eye movements during multiple fixation search. *J. Vis.*, 4(8):165–165, 8 2004.

[11] J. M. Henderson, J. R. Brockmole, M. S. Castelhano, and M. Mack. Visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. V.

Gompel, M. H. Fischer, W. S. Murray, and R. L. Hill, editors, *Eye Movements*, pages 537 – 562. Elsevier, Oxford, 2007.

[12] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40:1489–1506, 2000.

[13] J. Jovancevic, B. Sullivan, and M. Hayhoe. Control of attention and gaze in complex environments. *Journal of Vision*, 6:1431–1450, 2006.

[14] M. F. Land. Predictable eye-head coordination during driving. *Nature*, 359:318–320, September 1992.

[15] S. J. Landry, T. B. Sheridan, and Y. M. Yufik. A methodology for studying cognitive groupings in a target-tracking task. *IEEE Transactions on Intelligent Transportation Systems*, 2(2), June 2001.

[16] S. E. Lee, , C. B. Olsen, and W. W. Wierwille. A comprehensive examination of naturalistic lane changes. *Report DOT HS 809702, NHTSA, U.S. Department of Transportation*, March 2004.

[17] Y. C. Lee and J. D. L. amd L. N. Boyle. Visual attention in driving: the effects of cognitive load and visual disruption. *Human Factors*, 49(4):721–733, 2007.

[18] J. Levy, H. Pashler, and E. Boer. Central interference in driving: Is there any stopping the psychological refractory period? *Psychological Science*, 17(3):228–235, 2006.

[19] Y. Matsumoto, T. Ogasawara, and A. Zelinsky. Behavior recognition based on head pose and gaze direction measurement. *Intelligent Robots and Systems, 2000. (IROS 2000). Proceedings. 2000 IEEE/RSJ International Conference on*, 3:2127–2132 vol.3, 2000.

[20] J. C. McCall and M. M. Trivedi. Driver behavior and situation aware brake assistance for intelligent vehicles. *Proceedings of the IEEE*, 95(2):374–387, Feb 2007.

[21] J. C. McCall, D. Wipf, M. M. Trivedi, and B. Rao. Lane change intent analysis using robust operators and sparse bayesian learning. *IEEE Transactions on Intelligent Transportation Systems*, Sept. 2007.

[22] P. Morasso, G. Sandini, V. Tagliasco, and R. Zaccaria. Control strategies in the eye-head coordination system. *Systems, Man and Cybernetics, IEEE Transactions on*, 7(9):639–651, Sept. 1977.

[23] E. Murphy-Chutorian, A. Doshi, and M. M. Trivedi. Head pose estimation for driver assistance systems: A robust algorithm and experimental evaluation. *IEEE International Transportation Systems Conference*, Sept 2007.

[24] E. Murphy-Chutorian and M. M. Trivedi. Head pose estimation in computer vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 31(4):607–626, April 2009.

[25] V. Navalpakkam and L. Itti. Modeling the influence of task on attention. *Vision Research*, 45:205–231, 2005.

[26] N. Oliver, B. Rosario, and A. Pentland. A bayesian computer vision system for modeling human interactions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):831–843, Aug 2000.

[27] M. Peden, R. Scurfield, D. Sleet, D. Mohan, A. A. Hyder, E. Jarawan, and C. Mathers, editors. *World report on road traffic injury prevention*. World Health Organization, Geneva, 2004.

[28] J. Pelz, M. Hayhoe, and R. Loeber. The coordination of eye, head, and hand movements in a natural task. *Exp Brain Res.*, 139(3):266–277, August 2001.

[29] A. Pentland and A. Liu. Modeling and prediction of human behavior. *Neural Computation*, 11:229–242, 1999.

[30] R. J. Peters and L. Itti. Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention. *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.

[31] M. A. Recarte and L. M. Nunes. Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied*, 6(1):31–43, 2000.

[32] C. A. Rothkopf, D. Ballard, and M. Hayhoe. Task and context determine where you look. *Journal of Vision*, 7(14):1–20, 2007.

[33] M. S. Ryoo and J. K. Aggarwal. Hierarchical recognition of human activities interacting with objects. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:1–8, 2007.

[34] M. E. Tipping and A. C. Faul. Fast marginal likelihood maximisation for sparse bayesian models. In C. M. Bishop and B. J. Frey, editors, *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, Jan 3-6 2003.

[35] M. M. Trivedi and S. Y. Cheng. Holistic sensing and active displays for intelligent driver support systems. *IEEE Computer*, May 2007.

[36] M. M. Trivedi, T. Gandhi, and J. C. McCall. Looking in and looking out of a vehicle: Computer vision-based enhanced vehicle safety. *IEEE Transactions on Intelligent Transportation Systems*, January 2007.

[37] P. Viola and M. J. Jones. Rapid object detection using a boosted cascade of simple features. *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[38] J. Wu and M. M. Trivedi. Simultaneous eye tracking and blink detection with interactive particle filters. *EURASIP Journal on Advances in Signal Processing*, Oct 2007.

[39] S. Yantis. Control of visual attention. *in Attention, ed. H. Pashler. 13-74 Psychology Press, Hove, UK.*, 1998.

[40] W. H. Zangemeister and L. Stark. Types of gaze movement: Variable interactions of eye and head movements. *Experimental Neurology*, (77):563–577, 1982.

[41] G. J. Zelinsky, W. Zhang, B. Yu, X. Chen, and D. Samaras. The role of top-down and bottom-up processes in guiding eye movements during visual search. *Neural Information Processing Systems Conference*, 2005.