Investigating the Relationships Between Gaze Patterns, Dynamic Vehicle Surround Analysis, and Driver Intentions

Anup Doshi and Mohan Trivedi

Abstract— **Recent advances in driver behavior analysis for Active Safety have led to the ability to reliably predict certain driver intentions. Specifically, researchers have developed Advanced Driver Assistance Systems that produce an estimate of a driver's intention to change lanes, make an intersection turn, or brake, several seconds before the act itself. One integral feature in these systems is the analysis of driver visual search prior to a maneuver, using head pose and eye gaze as a proxy to determine focus of attention. However it is not clear whether visual distractions during a goal-oriented visual search could change the driver's behavior and thereby cause a degradation in the performance of the behavior analysis systems. In this paper we aim to determine whether it is feasible to use computer vision to determine whether a driver's visual search was affected by an external stimulus. A holistic ethnographic driving dataset is used as a basis to generate a motion-based visual saliency map of the scene. This map is correlated with predetermined eye gaze data in situations where a driver intends to change lanes. Results demonstrate the capability of this methodology to improve driver attention and behavior estimation, as well as intent prediction.**

I. Introduction

Advanced driver assistance systems should save many lives by aiding drivers to make prompt, safe decisions about driving maneuvers. To counter the effect of inattention, ADAS's could be designed to provide the driver ample warning time to impending dangerous situations, and even assist the driver in reacting appropriately. Successful design of these systems will rely heavily on getting the driver to accept the recommendations of the system, mainly by improving design to minimize false positives. The detection of driver intentions will become a crucial feature of future ADASs, therefore the focus of this paper is on improving the capabilities of driver intention inference systems.

In this paper we develop a framework to use computer vision systems to determine whether an external event potentially influenced a driver's gaze shift. Using ethnographic driving data, we collect video looking at the driver as well as an omnidirectional video looking around the vehicle. These data are processed using motion analysis to produce an estimate of whether there is a salient, or distracting, object in the area where the driver is looking. This information is then used to inform whether the gaze shift was due more to the salient object distracting the driver, as opposed to a task-oriented visual search. The intent of the driver could thus be more accurately inferred.

Authors with LISA: Laboratory for Safe and Intelligent Automobiles, University of California, San Diego, http://cvrr.ucsd.edu/lisa.
{andoshi,mtrivedi}@ucsd.edu

A. Motivations

Recent research has supported the incorporation of sensors looking inside the vehicle to observe driver behavior and potentially infer driver intent [1], [2]. Prior research has determined the more useful driver cues for distinguishing intent. Head motion has proved to be the most useful cue, whereas eye gaze tends to be noisier and less useful [3]. The head motion cue, along with data coming from the vehicle and from lane position information, can be used as inputs to a Bayesian learning algorithm which in turn is consistently able to predict the intention of a driver. Such a system has been trained on maneuvers including lane changes, intersection turns, and brake assistance systems [4]–[6].

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 we try to measure head motion given that the driver is likely paying attention to whatever they are looking at, in whichever direction they are looking. We then infer 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.

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 [7]–[11]. These stimuli may include visual distractions, which could pop up in a scene and thereby attract the attention of the observer. This research has concluded that for the most part any visual search is guided by some combination of a goal-driven and stimulus-driven approach, depending on the situation [12].

In the particular context of driving, several studies have shown that gaze and attention are directed in areas associated with particular goals and tasks [8], [13]. However it is also clear that distractions pose a significant risk to driving safety [14], [15], as there are important occasions when gaze and attention are directed due to external stimuli. In order to improve the performance of intent prediction systems, it would be useful to know whether a gaze shift is attributable more to irrelevant stimuli than to a specific goal. The intent prediction system could thereby acknowledge and potentially discard behavioral cues arising due to these distractions.

Additionally, the framework proposed here could lead to the first set of systematic examinations of the relationships between driver visual search and driver intentions. These

978-1-4244-3504-3/09/$25.00 ©2009 IEEE

experiments could answer questions such as how and why drivers change goals and behaviors in response to the environment surrounding them.

The remainder of the paper is organized as follows. Section II describes prior works in the field. In Section III we discuss the data processing steps, and Section IV includes the results. Finally we conclude and provide some future directions in Section V.

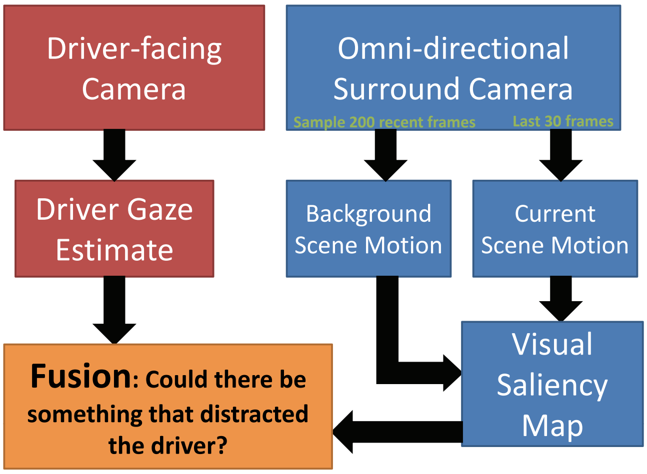

Fig. 1. Data flow architecture for the proposed system. The ultimate goal is to automatically achieve a determination of the cause for a driver's gaze shift.

II. Prior Works

A number of studies have focused on the question of determining what the driver is looking at. However, we are not as interested specifically in what the driver is looking at, but rather whether there is a distraction or unusual occurrence in the region where the driver is looking. To this end we review studies concerning determining driver gaze as well as scene saliency and dynamic environment analysis.

A. Gaze detection and Driving

To detect where the driver is looking, robust monocular in-vehicle head pose estimation systems have been developed [16]–[18], though it may be argued that head pose is not a sufficient estimate of true gaze. More precise gaze estimates can be derived from eye gaze detectors [19]. NHTSA has most recently conducted studies of Driver Workload Metrics [14], including eye gaze as a proxy for driver workload.

There have been a few studies that examine gaze behavior in the specific context of lane changes. According to Tijerina et al. [20], there are specific eye glance patterns which take place in the period before a lane change. Prior to lane changes, there were between 50-80% probabilities of glancing at a mirror. Mourant and Donohue observed that lengthy blind spot checks occured only in conjunction with lane change maneuvers; in lane keeping situations no such checks were performed by the drivers [21]. The experiments of Land [13] obtain results in a real automotive setting.

Several studies have shown that fatigue [14], [22], traffic [23], and other cognitive and visual distractions [15] may have an effect on driver behavior. Cognitive distractions involve mental tasks that increase the workload of the driver without necessarily adding visual clutter, whereas visual distractions draw the driver's focus of attention away from the road. Recarte and Nunes [15] measured the number of glances to the mirror during a lane change, and found cognitive distractions decrease the glance durations from 3-4% to less than 1%, and visual distractions decrease the glance durations from 1.4% to 0.2-0.4%. This result is well-aligned with more recent results indicating the limitations of drivers' multi-tasking abilities [24]. These studies show that distractions do have a significant effect on the gaze patterns of drivers.

Huang et al. [25] did some work to identify the gaze position of the driver, and determine what objects are being observed. That study used a single omni-directional camera to observe the driver and the scene, and was limited by resolution and contrast issues. In the following study we use separate cameras for looking in and looking out of the vehicle.

B. Surround Analysis and Saliency

The analysis of gaze patterns can be aided by developing a robust saliency detector which can determine the relative attractiveness of objects in a scene. The potential structure of these saliency maps vary based on the motivation and context of the scene. A primitive saliency map could include features based on intensity, color, and orientation. Itti extended this saliency map by examining the eye glance patterns of human subjects on several scenes and building up a prior of that particular type of scene [7]. Recent studies, however, show that such "bottom-up" saliency maps can not explain the fixations of goal-oriented observers [10], [11]. More recent versions of saliency maps have thus incorporated top-down goals [8], [9], however much of the development of all of these have been motivated by images with relatively static backgrounds.

Ultimately it is the interaction between an observer's goals, and the salient properties of the scene, that guide the observer's attention [12], [26]. Itti found that motion cues are much stronger predictors of gaze changes than any other cue in complex scenes [27]. Therefore given that we are in the specific context of driving, with highly dynamic background scenes, we ultimately choose to use motion-based features to build up the saliency map, as detailed below. There have been similar motion-based approaches to scene analysis using omnidirectional cameras, in order to identify and track interesting objects in a scene [28]–[30].

III. Data Collection and Analysis

The data flow overview developed here can be seen in Figure 1. The driver view images are processed to produce an estimate of the gaze location, as described in Section III-A. In parallel, Section III-B describes how the surround images are analyzed using optical flow and background motion is subtracted to produce a saliency map. The final fusion and decision process is detailed in Section III-C.

A. *Gaze Estimation*

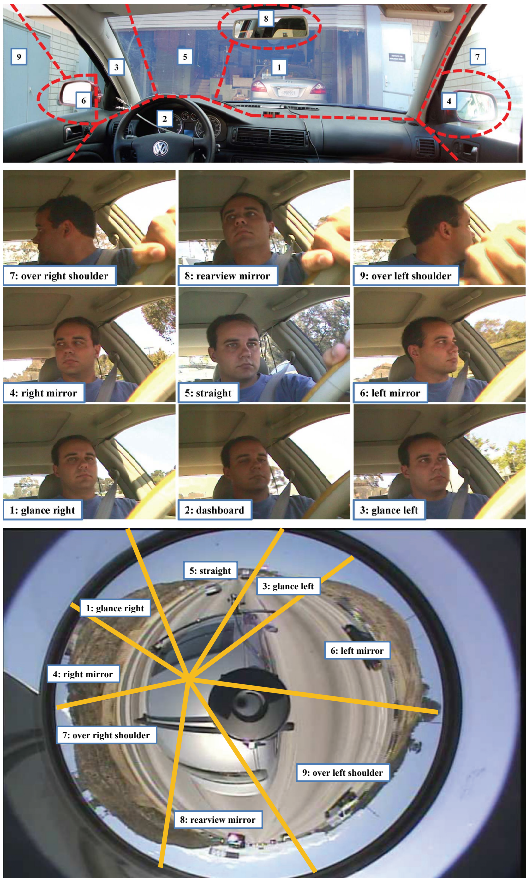

Fig. 2. (Top) Approximate distribution of eye gaze location classifications. (Middle) Samples from dataset showing corresponding eye gaze locations. (Bottom) Approximate location of glance directions superimposed on the surround map. Note that the omnidirectional image is inverted horizontally, so that the right side of the image is the driver's left.

The driver-facing monocular camera was mounted above the radio controls, looking at an angle toward the driver. This angle turned out to be too obtuse for monocular eye gaze estimators such as [31] to work reliably. Ideally, a properly designed stereo or monocular eye gaze system could provide robust data. Head pose estimators [16] could also be used to determine gaze direction, though not as precisely.

In fact there have been real-world studies that relied on automatically detecting eye gaze as a proxy for fatigue, but their results were limited due to robustness issues, especially with regards to occlusions from sunglasses and harsh lighting conditions [32]–[35]. Several studies that achieved promising results using just head motion as a cue for behavior prediction [4]–[6], [36].

In order to approximate an ideal case and demonstrate the feasibility of the system, the gaze data was manually reduced. The procedure followed was similar to those followed in the NHTSA lane-change and workload experiments [14], [37], to produce output that a real-world eye gaze tracker would output in an optimal setting. Nine different gaze locations were derived from the procedure described in [14] as relevant to the task at hand. Sample images from each of these cases can be seen in Figure 2.

Once the gaze direction is determined, it becomes necessary to calibrate the gaze direction with a section of the environment in the outside camera. The approximate regions corresponding to each gaze direction can be seen in the bottom of Figure 2.

B. *Surround Analysis*

Fig. 3. Omni-directional Camera-based Surround of vehicle. Superimposed is the "Background Motion Map": the optical flow map of the average background motion in the scene, captured using optical flow of several hundred frames sampled over the previous 10 seconds of data.

Visual saliency maps are produced by extracting useful and pertinent features of the surrounding environment that may attract the driver's attention. The development of several different kinds of such saliency maps has been detailed above, and here we extend that notion to work for arbitrary cameras in a moving vehicle using motion-based features.

As discussed above, there have been a number of methodologies introduced to develop saliency maps of visual fields. Given that we have the advantage of knowing the context has to do with driving, we can develop a saliency model that is more directly related to driving. Indeed, given that a driver is most often looking forward at the road, a visual distraction will most likely pop up in the peripheral vision of the driver. Specifically, the peripheral field is the region that subtends between $10\,\mathrm{deg}$ and $180\,\mathrm{deg}$ on either side of the human visual field. Research has shown that the peripheral visual field of human eyes is most sensitive to motion cues [38]. Therefore in this system we generate a saliency map of the scene based primarily on these motion cues.

In order to find interesting motion in a dynamic scene where nearly everything is moving, it is imperative find things that are moving in places where normally things move in another direction. In order to simplify the problem, here we deal only with highway driving, as city driving involving stops and drastic speed changes would introduce greater obstacles. Generally on highways there may be cars or other objects that are moving at the same speed and direction as the ego-vehicle, but most interesting motion would be of those objects which are moving with a different velocity than is "normal" in that area of the scene.

To detect motion, we use the Lucas-Kanade dense optical flow algorithm, comparing the previous frame with the current frame. The data is a 720x480 resolution Omnidirectional video at 30 frames per second, so at the relatively high highway speeds the inter-frame motion is on the order of several pixels.

To account for the "normal" motion in the scene, we build up a "background" motion map. The optical flow vectors for each pixel are accumulated over several hundred frames, sampled over the last ten seconds of data. The average of these detections becomes our "Background Motion Map". Figure 3 shows an example of the background motion vectors in the scene, superimposed on the scene itself.

Then in order to detect "foreground" motion that could be of interest to the driver, we can collect the most current motion vectors and subtract out the Background Motion Map. Examples of this foreground motion detection can be seen in Figures 5 and 6. However the instantaneous motion tends to be noisy, so to be more robust we can average the motion vectors over the previous second of data, and then subtract the background motion to obtain our foreground motion map. Examples can be seen in Figure 4.

The "saliency maps" on the bottom of Figure 4 are generated by taking the magnitude of the foreground motion map. These maps thereby display the more interesting scene elements as ones which have more motion as compared to the typical background motion.

C. Fusion and Decision

Given the location of the gaze direction, along with the saliency map, it is straightforward to determine whether there was a salient object in the driver's view. The right column of Figures 5 and 6 show this fusion step.

The decisions or conclusions to be drawn given this information can have significant implications. Given a particular gaze, if there is no salient object in the region, then it is much more likely that the driver's attention is motivated by a goal the driver had in mind.

For certain glances, there are indeed salient objects in the region. Figure 6 shows an example of this. For these cases the gaze changes may possibly be due to distractions caused by that salient object. The gaze shifts may also be due to driver's goals as above, but given that the driver saw the salient object, their goals may also change. Either way, if there is a salient object, the intent of the driver can not be as easily inferred as if there is no salient object.

IV. Results

The output of the surround analysis step can be seen in Figure 4. Results of the overall system can be seen in Figures 5 and 6. These sequences demonstrate the potential improvements to driver behavior recognition and intent recognition that could arise from the system.

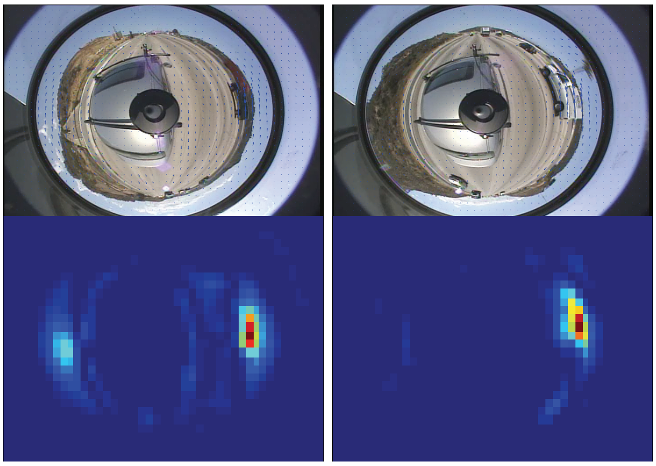

Fig. 4. (Top) Average motion in the scene over the previous second, after discarding any normal "background" scene motion. (Bottom) Saliency map of the scene, displaying the motion map from above in terms of magnitude of motion.

In the sequence of frames in Figure 5 below, the driver is clearly looking over their right shoulder, as the eye gaze analysis dictates. However there are no cars in that lane, nor are there any other salient objects in the scene. The saliency map produced from the surround analysis shows that there are only salient objects on the other side of the scene (on the driver's left). We may thereby conclude, that the gaze shift which occurred in that window of time was associated with a specific goal of the driver. In this case, a highway situation, that goal would most likely be a lane change, and so an intent prediction system could more confidently predict the upcoming lane change.

The next sequence in Figure 6 allows us to draw a different conclusion. Here the driver looks to his left, approximately toward his left mirror. In this case the driver could potentially be looking in the mirror to see if there are any cars in the left blind spot area. However we also note that the saliency map produces a salient object in the region of interest. This vehicle is traveling at a greater speed than the ego-vehicle, and so may attract some attention. Therefore we may hesitate to conclude that the driver's visual search is due to an intention to change lanes.

V. Concluding Remarks

We have proposed a system designed to fuse the computer vision-based detections of gaze patterns and environmental motion saliency maps. We are thus able to determine the effects of the surrounding environment on the attention of the driver. The implications could have significant effects on the design of driver behavior analysis and intent prediction systems.

Future work includes testing these results using an automatic eye gaze analysis system. Additionally, the saliency map could be compared with other types of saliency maps, including those based on feature detectors and object detectors. Improvements could be made to the saliency map to

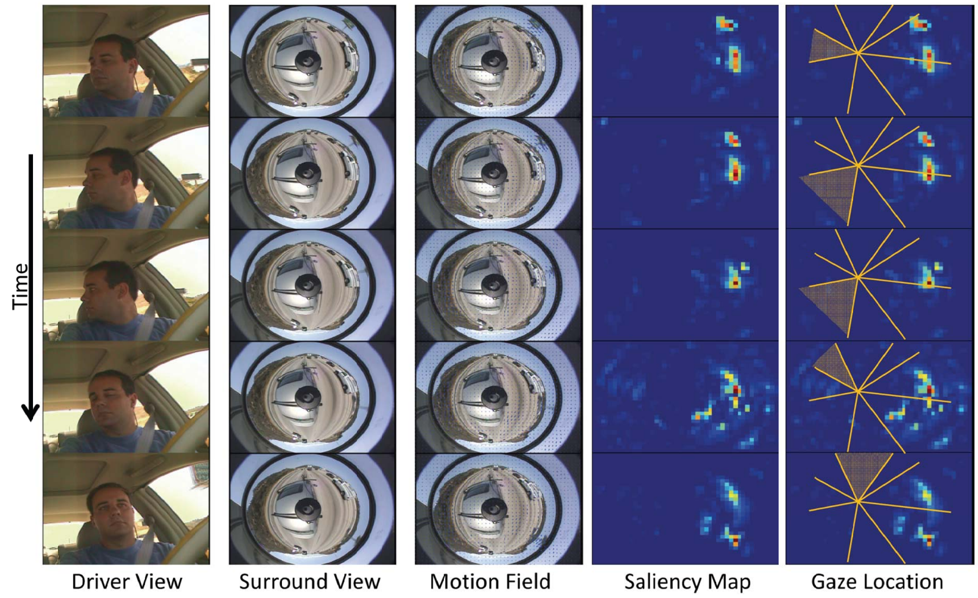

Fig. 5. Example sequence (1). The leftmost column shows the driver looking over his right shoulder as the sequence progresses, which corresponds to the left side of the Omni image. Using the surround analysis algorithm we determine the "foreground" motion vectors in the surrounding scene, and then produce a saliency map. Note the vehicle in the scene is detected in the saliency map, but is actually on the driver's left side. The final column superimposes the glance direction, and we discover that there are no salient objects which caught the driver's attention.

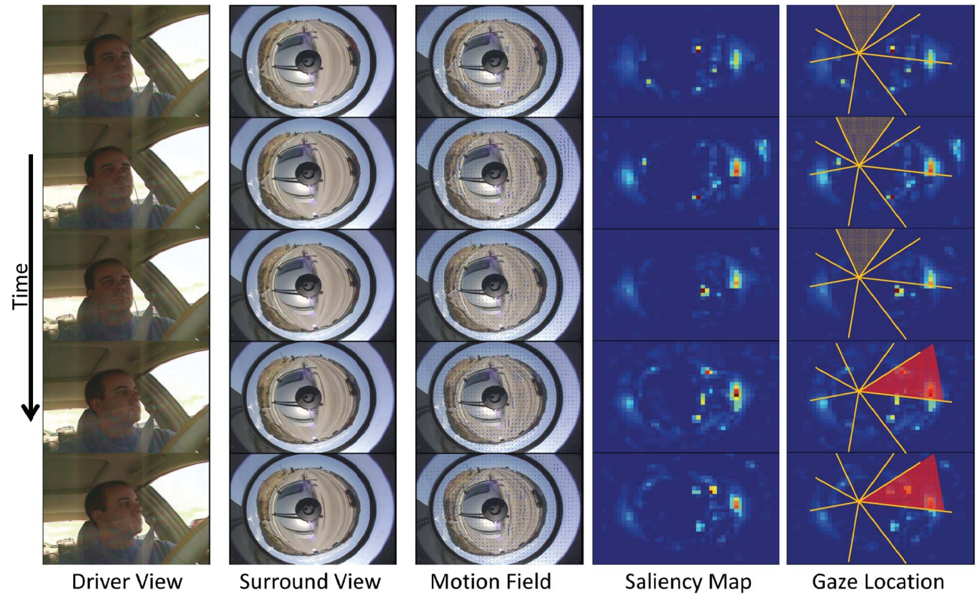

Fig. 6. Example sequence (2). This time the driver is looking straight and then glances to his left. As seen in the surround camera, there is indeed a vehicle in this region of the environment. The saliency map detects that vehicle, allowing the system to conclude that in the last two frames, the attention of the driver may have been distracted by that vehicle.

work in more dense traffic situations and in urban environments.

Additionally, the framework proposed is general enough to be able to use other hardware setups. Potential commercially deployable systems may include a camera in the dashboard looking at the driver, and a camera in the rearview mirror looking out at the road in front. Future work may include further analysis of this framework in this and other relevant settings, as well as in more environments.

Acknowledgements

The authors would like to thank Volkswagen and the UC-Discovery DiMI Grant for their support. We appreciate the assistance of Dr. Joel McCall in collecting data and our colleagues in LISA for their valuable advice and assistance.

References

[1] 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.

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

[3] 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.

[4] 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*, vol. 5, no. 4, pp. 28–37, October-December 2006.

[5] 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.

[6] J. C. McCall and M. M. Trivedi, "Driver behavior and situation aware brake assistance for intelligent vehicles," *Proceedings of the IEEE*, vol. 95, no. 2, pp. 374–387, Feb 2007.

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

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

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

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

[11] C. A. Rothkopf, D. Ballard, and M. Hayhoe, "Task and context determine where you look," *Journal of Vision*, vol. 7, no. 14, pp. 1–20, 2007.

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

[13] M. F. Land, "Predictable eye-head coordination during driving," *Nature*, vol. 359, pp. 318–320, September 1992.

[14] L. Angell, J. Auflick, P. 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.

[15] M. A. Recarte and L. M. Nunes, "Effects of verbal and spatial-imagery tasks on eye fixations while driving," *Journal of Experimental Psychology: Applied*, vol. 6, no. 1, pp. 31–43, 2000.

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

[17] S. Baker, I. Matthews, J. Xiao, R. Gross, T. Ishikawa, and T. Kanade, "Real-time non-rigid driver head tracking for driver mental state estimation," *CMU-RI-TR-04-10*, Feb 2004.

[18] Y. Zhu and K. Fujimara, "Headpose estimation for driver monitoring," *IEEE Intelligent Vehicles Symposium*, 2004.

[19] M. Smith and H. Zhang, "Safety vehicles using adaptive interface technology (task 8): A literature review of intent inference," *SAVE-IT Phase 1 Final Report, http://www.volpe.dot.gov/hf/roadway/saveit/*, 2004.

[20] L. Tijerina, W. R. Garrott, D. Stoltzfus, and E. Parmer, "Eye glance behavior of van and passenger car drivers during lane change decision phase," *Transportation Research Record*, vol. 1937, pp. 37–43, 2005.

[21] R. Mourant and R. Donahue, "Mirror sampling characteristics of drivers," *Society of Automotive Engineers*, no. 740964, 1974.

[22] R. Hanowski, W. W. Wierwille, S. A. Garness, and T. A. Dingus, "A field evaluation of safety issues in local/short haul trucking," *Proceedings of the IEA 2000/HFES 2000 Congress*, 2000.

[23] G. Robinson, D. Erickson, G. Thurston, and R. Clark, "Visual search by automobile drivers," *Human Factors*, vol. 14, no. 4, pp. 315–323, 1972.

[24] J. Levy, H. Pashler, and E. Boer, "Central interference in driving: Is there any stopping the psychological refractory period?" *Psychological Science*, vol. 17, no. 3, pp. 228–235, 2006.

[25] K. S. Huang, M. M. Trivedi, and T. Gandhi, "Driver's view and vehicle surround estimation using omnidirectional video stream," *Proc. IEEE Intelligent Vehicles Symposium*, pp. 444–449, 2003.

[26] 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*, 2005.

[27] L. Itti, "Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes," *Visual Cognition*, vol. 12, no. 6, pp. 1093–1123, 2005.

[28] T. Gandhi and M. M. Trivedi, "Video based surround vehicle detection, classification and logging from moving platforms: Issues and approaches," *Proc. IEEE Intelligent Vehicles Symposium*, 2007.

[29] ——, "Vehicle surround capture: Survey of techniques and a novel omni video based approach for dynamic panoramic surround maps," *IEEE Transactions on Intelligent Transportation Systems*, Sept. 2006.

[30] ——, "Parametric ego-motion estimation for vehicle surround analysis using an omnidirectional camera," *Machine Vision and Applications*, vol. 16, no. 2, pp. 85–95, Feb. 2005.

[31] 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.

[32] X. Liu, F. Xu, and K. Fujimura, "Real-time eye detection and tracking for driver observation under various light conditions." in *Proceedings of IEEE Intelligent Vehicles Symposium*, June 2002, pp. 18–20.

[33] R. Smith, M. Shah, and N. da Vitoria Lobo, "Determining driver visual attention with one camera," *IEEE Transactions on Intelligent Transportation Systems*, vol. 4, pp. 205–218, December 2003.

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

[35] L. Bergasa, J. Nuevo, M. Sotelo, R. Barea, and M. Lopez, "Real-time system for monitoring driver vigilance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, pp. 63–77, March 2006.

[36] S. Y. Cheng, S. Park, and M. M. Trivedi, "Multi-spectral and multi-perspective video arrays for driver body tracking and activity analysis," *Computer Vision and Image Understanding* in press, 2007, doi:10.1016/j.cviu.2006.08.010.

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

[38] M. Rizzo and I. L. Kellison, "Eyes, Brains, and Autos," *Arch Ophthalmol*, vol. 122, pp. 641–645, Apr. 2004.