2014 IEEE Intelligent Vehicles Symposium (IV)
June 8-11, 2014. Dearborn, Michigan, USA

Frame-Rate Vehicle Detection within the Attentional Visual Area of Drivers

S.M. Zabihi[1] S.S. Beauchemin[1] E.A.M. de Medeiros[2] M.A. Bauer[1]

Abstract— **This contribution consists of a frame-rate, vision-based, on-board method which detects vehicles within the attentional visual area of the driver. The method herein uses the 3D absolute gaze point of the driver obtained through the combined use of a front-view stereo imaging system and a non-contact 3D gaze tracker, alongside hypothesis-generation reducing techniques for vehicular detection such as horizon line detection. Trained AdaBoost classifiers are used in the detection process. This technique is the first of its kind in that it identifies vehicles the driver is most likely to be aware of at any moment while in the act of driving.**

I. Introduction

It has been known for some time that driver ocular movements often provide information on the very next maneuver to be effected [1], such as in the case of negotiating a road bend [2]. Our research program focuses on the development of intelligent, Advanced Driving Assistance Systems (i-ADAS) which include the driver as a pivotal element of driving error prediction and correction.

A. Context

The underlying principle underpinning this research is to determine to what extent current maneuvers and ocular behavior are predictive of the next maneuver and also whether this predicted maneuver is consonant with the current driving context (other vehicles, pedestrians, signage, and lanes, among others). Our current vision for i-ADAS is that of intervening (by a switch to autonomous driving) only when the next predicted maneuver poses a measurable traffic risk. Vehicular control is then returned once the danger has abated and the driver is capable of correctly controlling the vehicle. Such an i-ADAS, once realized, would certainly significantly reduce vehicle accidents, injuries, and fatalities, without removing the driver from the act of commuting.

The complexities involved in such an endeavor require many stages of research, such as the study of driver behavior as it pertains to ocular movements and current maneuvers, the precise detection and understanding of traffic context surrounding the vehicle, and a correct assessment of objects that are perceived (or not) by the driver. This contribution relates to the latter, as it consists of an attempt at detecting vehicles located in the attentional visual field of the driver.

[1]Department of Computer Science, Western University, London, ON, N6A 5B7, Canada
[2]Key Factors Inc. 417 Beaconsfield Blvd. Beaconsfield, QC, H9W 4B4, Canada

B. Background

More than 30,000 people die in vehicle-related accidents in the United States every year [3]. Current Driver Assistance Systems (DAS) such as Electronic Stability Control have reduced these grim numbers over the recent years. However much more remains to be accomplished since driver actions, maneuvers, and ocular behavior as they relate to the environment are usually not included in the retroactive mechanisms composing DAS. Given that most vehicle accidents are rooted in human error, the next steps in ADAS research are clearly laid out. One crucial task is to automatically and reliably detect relevant objects in the environment which the driver may or may not have seen, as a starting point of a driver-aware assistance system.

Many vision-based algorithms have been developed for the purpose of detecting vehicles at frame rate in recent years, but many such techniques lack the required accuracy and speed to be of any practical use. Reducing the search space for Hypothesis Generation (HG) and extracting only the relevant information may assist in reaching higher efficiency. The most relevant information concerning what a driver is looking at comes from the 3D gazing point of the driver in absolute coordinates and what type of objects are located around it [4]. In this contribution our interest focuses on the vehicles the driver is most likely to be aware of at any given time. The experiments were conducted with an improved version of the RoadLAB experimental vehicle described in [5].

The remainder of this contribution is organized as follows: Section II describes the current literature in the area of vehicle detection. The proposed technique is presented in Section III. Results and evaluations are given in Section IV. Section V concludes this work.

II. Literature Survey

Vehicle detection has been and continues to be an important area of research in Computer Vision. Statistical methods such as SVM [6], [7], PCA [8] or Neural Networks [9] constituted the main body of early approaches. However, these did not achieve the levels of efficiency required for realistic use.

In 2001, Viola and Jones proposed a powerful method for rapid object detection based on training a sequence of classifiers using Haar-like features [10]. The authors demonstrated high performance rates for their method by applying it to the problem of facial recognition [11]. This approach became widely used in areas including vehicle

978-1-4799-3637-3/14/$31.00 ©2014 Crown

detection [12], and many researchers have focused on its variants in recent years [13].

In general, any detection problem is formulated in two stages: Hypothesis Generation (HG) and Hypothesis Verification (HV). Various strategies have been used for HG. Some authors used shadows cast on the pavement by the presence of vehicles [14], while others exploited edge maps [15], vehicle symmetry [16], and so on, as cues to finding vehicles in sets of images. Reducing the range of probable vehicle positions in HG is of utmost importance to achieve decent frame rates in the vehicle detection process. In the case of HV, most techniques apply AdaBoost as a strong classifier to verify the existence of vehicles in the the image areas identified as likely candidates by the hypothesis generation mechanism. For instance, Khammari *et al.* used a gradient based algorithm for hypothesis generation and AdaBoost classifier for candidate verification [17]. Alternatively, Song *et al.* generated candidate image regions by use of edge maps and properties of vehicle symmetry followed by texture analysis and AdaBoost for vehicle detection [18]. None of these approaches are attempting to determine which vehicles are in the attentional visual field of the driver. Rather, they attempt to identify all vehicles within image sequences.

III. Proposed Method

For reasons cited above, our interest lies in the detection of vehicles that a driver is most likely to be aware of at any moment. For this reason, we have selected two approaches developed in our laboratories and combined them into the technique we present herein. The first of our techniques, due to Kowsari *et al.* performs vehicle detection by using Hypothesis Generation techniques that significantly reduce the number of candidate image regions while preserving excellent detection rates [19]. The second technique we adopted in this contribution pertains to the identification of the 3D point of gaze in absolute 3D coordinates given in the frame of reference of the vehicle [20]. This technique is unique in that it avoids the parallax errors introduced by the combined usage of 3D gaze trackers with monocular scene cameras. As expected, adjoining these two techniques allows for even more stringent HG strategies since we are only interested in vehicles around the driver point of gaze, thus increasing the nominal frame rate for vehicle detection. Our driving sequences for test purposes were recorded with the RoadLAB experimental vehicle, as depicted in Figure 1. Both the forward stereo scene system and the non-contact 3D gaze tracker operate in their own frames of reference. While we define that of the scene stereo system to be that of the vehicle, there is a need to transform the 3D driver gaze, expressed in the coordinates of the tracker, into that of the vehicle. Fortunately, the technique to perform this transformation has been established by Kowsari *et al.* [20] in our laboratories.

A. Hypothesis Generation

The objective of HG is to identify candidate image areas likely to contain vehicles, in order to disregard image areas

Fig. 1. *Vehicular instrumentation configuration.* **a) (top):** *In-vehicle non-contact infra-red binocular gaze tracking system and forward scene stereo imaging system mounted on the roof of the vehicle.* **b) (bottom):** *On-board software systems.*

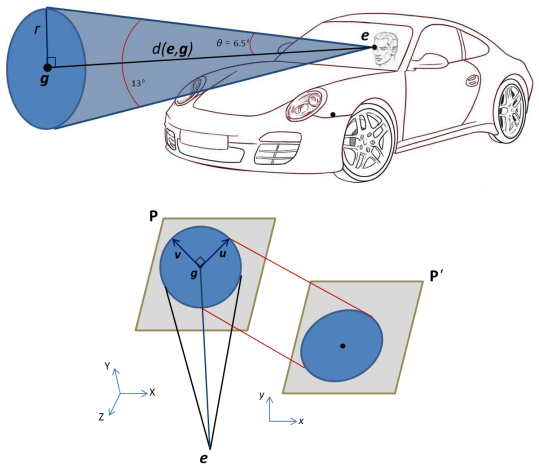

Fig. 2. **(top):** *A depiction of the driver attentional gaze cone. It is generally accepted that the radius of the cone is approximately* 13° *[21] and* **(bottom):** *reprojection of 3D attentional circle onto the 2D ellipsoid on image plane of the forward stereo scene system.*

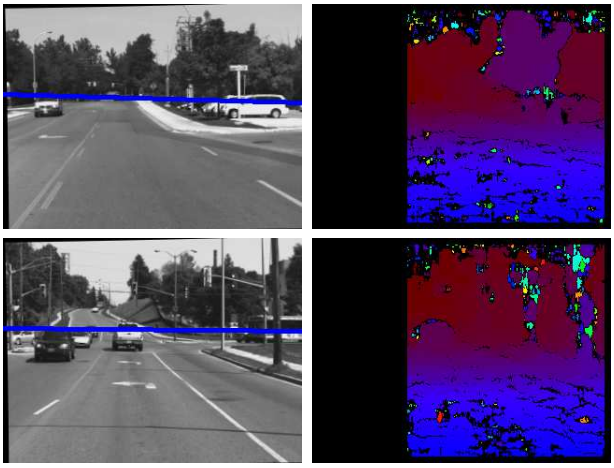

Fig. 3. **(left):** *Estimated horizon lines and* **(right):** *their corresponding depth maps.*

not likely to contain such objects (such as visible part of the sky, for instance). A considerable number of false positives may be removed by the suppression of image areas in the HG stage for which we know no vehicle can be present. One such powerful constraint applied to the HG process is the fact that vehicles will appear in imagery within a horizontal band centered around the horizon line (if we assume a relatively flat landscape). Furthermore, since our interest is focused on vehicles within an area around the 3D gaze of the driver, then the HG stage need only consider the intersection of the band around the horizon line and the 3D gaze area of the driver. This intersection constitutes a Region of Interest (RoI) within which the HG stage operates. Figure 2 depicts the attentional gaze cone of a driver. The radius of the circular gaze area depends on the distance between the fixated object and the driver.

The horizon line is approximated by intersecting the plane parallel to that of the ground plane and passing through the focal point with the image plane of the forward stereo scene system, and converting to image coordinates using stereo calibration parameters. The ground plane estimation is performed by grouping detected 3D ground points in areas of the imagery where no obstacle is known to be present. These points are then used in a robust plane-fitting computation to obtain the ground plane parameters (implementation details for both the ground plane and the horizon line are provided in [19]).

Once provided with an estimate of the horizon line, the first RoI is given by the horizontal band defined as

$$\text{RoI}_1 : f(x) = mx + b \pm \delta \tag{1}$$

where m and b specify the horizon line and δ determines the vertical span of the RoI around it. Vehicles are expected to be found within this image region. Figure 3 shows the horizon line for two different frames from a test sequence.

The second RoI of interest consists of the circle formed by the intersection of the visual cone of attention with the plane perpendicular to the 3D Line of Gaze (LoG) and containing the 3D Point of Gaze (PoG) of the driver. The hypothesis that visal attention operates as a spotlight within the visual field is corroborated by an number of studies [22]. Additionally, it seems reasonable to equate the span of this spotlight to $\pm 6.5^\circ$ from the pitch and yaw angles of the gaze direction, as it corresponds to the human central field of view [21], resulting in the attentional gaze cone depicted in Figure 2.

Once the 3D eye position $\mathbf{e} = (e_x, e_y, e_z)^T$ from which the LoG is emanating, and the 3D PoG $\mathbf{g} = (g_x, g_y, g_z)^T$ have been transformed into the frame of reference of the forward stereo scence system (as per [20]), the radius of the circular gaze area onto the plane perpendicular to the LoG and containing the PoG is obtained as

$$r = \tan(\theta) d(\mathbf{e}, \mathbf{g}) \tag{2}$$

where $\theta = 6.5^\circ$, and d is the Euclidean distance between $\mathbf{e}$ and $\mathbf{g}$, defined as

$$d(\mathbf{e}, \mathbf{g}) = \sqrt{(\mathbf{e} - \mathbf{g})^T (\mathbf{e} - \mathbf{g})} \tag{3}$$

At this point, the circle defined by the PoG and radius r and contained in the 3D plane perpendicular to the LoG is reprojected onto the image plane of the forward stereo scene system and delineates the 2D portion of the scene that falls onto the attentional visual area of the driver. Since this plane is generally not parallel to the image plane of the forward stereo scene system, the 3D attentional circle projects to an ellipsoid as illustrated in Figure 2. The 3D circle in parametric form can be written as

$$\mathbf{S}(\phi) = (X(\phi), Y(\phi), Z(\phi))^T = \mathbf{g} + r(\mathbf{u}\cos\phi + \mathbf{v}\sin\phi) \tag{4}$$

where $\mathbf{u} = (u_x, u_y, u_z)^T$ and $\mathbf{v} = (v_x, v_y, v_z)^T$ are two orthogonal unit vectors within the plane, and parameter ϕ varies from 0 to 2π. This circle is transformed into the reference frame of the stereo system as

$$\mathbf{S}'(\phi) = R^T(\mathbf{S}(\phi) - \mathbf{T}) \tag{5}$$

where R is a rotation matrix and $\mathbf{T}$ is a translation vector describing the rigid transformation from the reference frame of the stereo system to that of the eye-tracker. The image coordinates of the resulting ellipsoid are obtained as

$$\mathbf{s}'(\phi) = \frac{1}{Z'(\phi)} K \mathbf{S}'(\phi) \tag{6}$$

where

$$K = \begin{bmatrix} f_x & 0 & o_x \\ 0 & f_y & o_y \\ 0 & 0 & 1 \end{bmatrix}$$

is the matrix of intrinsic parameters related to the scene stereo system [20].

The second region of interest RoI_2 is thus defined as the image region contained within the 2D circle as it reprojects onto the image plane of the forward stereo scene system. The final RoI of interest for the HG mechanism is consequently defined as the intersection of RoI_1 and RoI_2. Given image region $\text{RoI} = \text{RoI}_1 \cap \text{RoI}_2$, the HG mechanism proceeds according to the algorithm given in [19]. Figure 4 displays

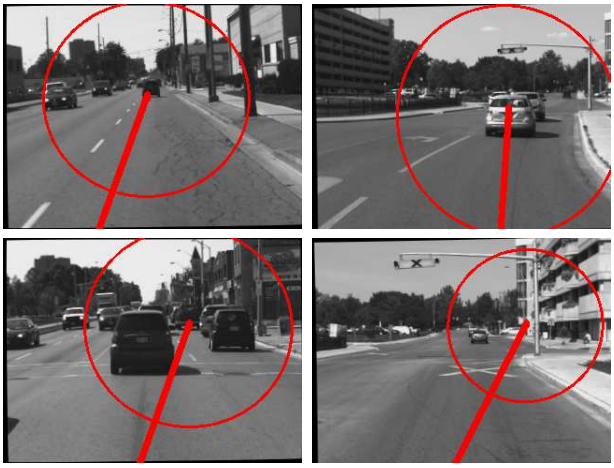

Fig. 4. *Displays of various LoGs, PoGs, and attentional gaze areas projected onto the forward stereo scene system of the vehicle.*

several Lines of Gaze (LoGs), Points of Gaze (PoG), and attentional visual areas for selected frames.

B. Hypothesis Verification

AdaBoost, introduced by Freund and Schapire [23], is a method for choosing and combining a set of weak classifiers to build a strong classifier. Adjoining the concept of the integral image as an efficient way of computing Haar-like features and cascaded AdaBoost, Viola and Jones introduced a powerful method for object recognition [10]. We adopted this approach for the HG stage and used four cascaded AdaBoost classifiers to discriminate positive from false-positive hypotheses. The Haar-training module from OpenCV is used to train our classifiers. Figure 5 shows various Haar-like features. We used in excess of two hundred vehicle images as positive samples for each classifier. For the negative examples, a set of more than five hundred images randomly downloaded from the Internet was used.

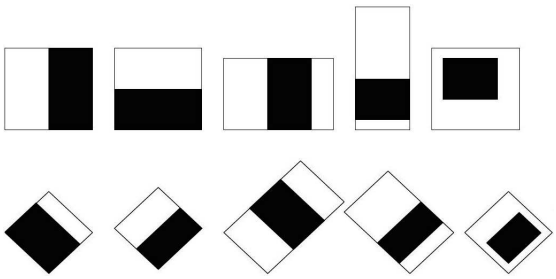

Fig. 5. *A depiction of Haar-like features.*

IV. Experimental Results

The proposed method was tested on 3,326 randomly selected frames, in which 5,751 vehicles in various views and lanes appeared. These were manually annotated for the purpose of evaluating our method. Figure 6 shows a small sample of our results where each green rectangular area within the RoI indicates a vehicular detection. Rates of detection and false positives per frame for various vehicle views are given Table I, where **FV**, **RV**, **FSV** & **BSV**, and **SV**, stand for Front View, Rear View, Front-Side and Back Side Views, and Side View, respectively, while **DR** and **FP/F** are Detection Rate and False Positives per Frame.

TABLE I

Detection rates and false positives per frame for different views and vehicles

	FV	RV	FSV & BSV	SV	ALL
DR	0.9877	0.9890	0.8401	0.8118	0.9867
FP/F	0.25	0.22	0.30	0.31	0.24

TABLE II

Comparison on frame rates, hit rates, and false positives for various methods

Authors	F/S	HR	FP
Chang and Cho [24]	5	99%	12%
Southall *et al.* [25]	16	99%	1.7%
Bergmiller *et al.* [26]		83.12%	16.7%
Sun *et al.* [27]	10	98.5%	2%
Alonso *et al.* [28]		92.63%	3.63%
Cheng *et al.* [29]	20	90%	10%
Kowsari *et al.* [19]	25	98.6%	13%
Our Method	30	98.9%	11%

Figure 7 shows the performance of our method measured using a Receiver Operating Characteristics (ROC) curve, characterizes the True Positive Ratio (TPR) versus False Positives (FP). TPR is obtained by dividing True Positives by the number of vehicles. Table II compares various vehicle detection methods with ours, where **F/S**, **HT**, and **FP** stand for Frames per Second, Hit Rate, and False Positives, respectively.

V. Conclusion

We presented a method for the detection of vehicles located within the attentional visual area of drivers as an initial attempt at identifying which visual stimuli elicit ocular responses from drivers. Our implementation is non-contact and operated on-board an experimental vehicle at frame rate (30Hz). We believe this contribution could easily be extended to include other visual stimuli drivers routinely encounter and attend to, such as pedestrians, cyclists, traffic lights, signs, and more. Ultimately, such an augmented technique will be integrated into i-ADAS as a means to identify objects drivers attend to, and those that they do not, at frame rate.

References

[1] F. Lethaus and J. Rataj, "Do eye movements reflect driving manoeuvres?" *IET Intelligent Transportation Systems*, vol. 1, no. 3, pp. 199–204, 2007.

[2] M. Land and J. Horwood, "Which parts of the road guide steering?" *Nature*, vol. 377, pp. 339–340, 1995.

[3] National Highway Safety Administration, "Traffic safety facts: 2010 data," U.S. Department of Transportation, Tech. Rep. DOT HS 811 625, 2012.

[4] E. Hartmann, "Driver vision requirements," Society of Automotive Engineers, Hillsdale, NJ, Tech. Rep. 700392, 1970.

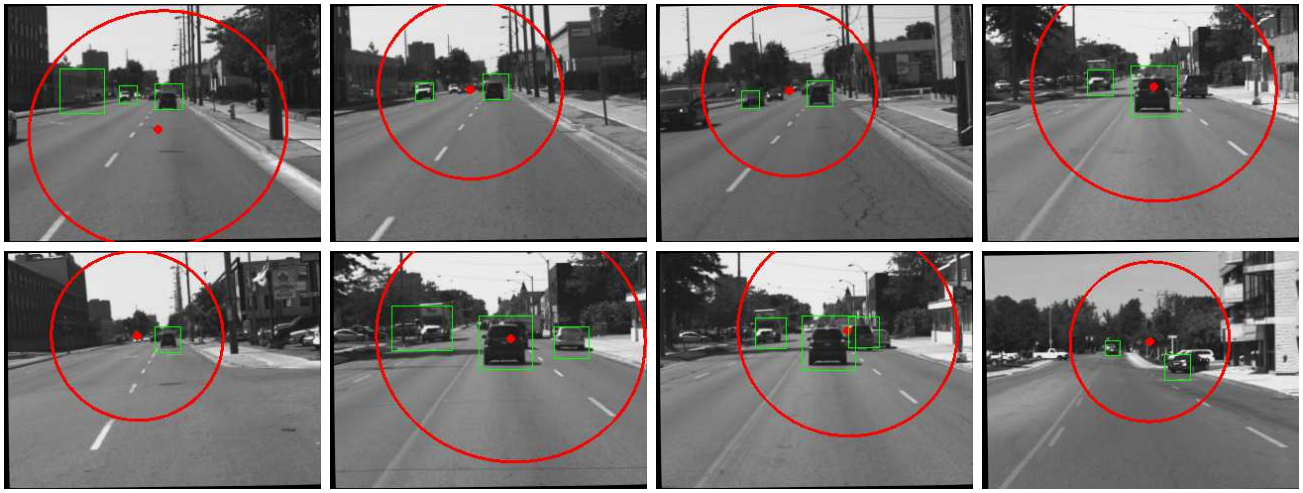

Fig. 6. *Detection results within the RoI, including false positives and negatives.*

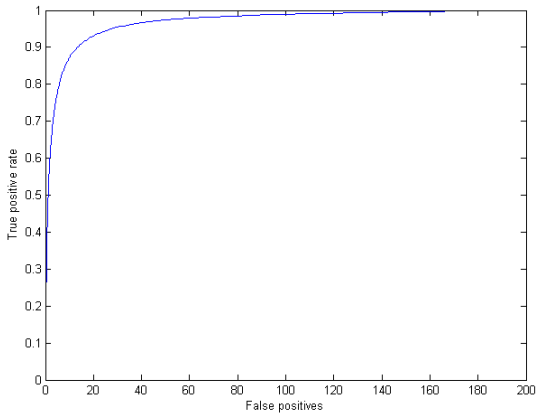

Fig. 7. *ROC curve obtained from experiments.*

[5] S. Beauchemin, M. Bauer, T. Kowsari, and J. Cho, "Portable and scalable vision-based vehicular instrumentation for the analysis of driver intentionality," *IEEE Trans. on Instr. and Meas.*, vol. 61, no. 2, pp. 391–401, 2012.

[6] C. Papageorgiou and T. Poggio, "A trainable system for object detection," *IJCV*, vol. 38, no. 1, pp. 15–33, 2000.

[7] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection using gabor filters and support vector machines," in *Int. Conf. on Digit. Sig. Proc.*, vol. 2, 2002, pp. 1019–1022.

[8] J. Wu and X. Zhang, "A pca classifier and its application in vehicle detection," in *Int. J. Conf. on Neural Networks*, vol. 1, 2002, pp. 600–604.

[9] N. Matthews, P. An, D. Charnley, and C. Harris, "Vehicle detection and recognition in greyscale imagery," *Control Engineering Practice*, vol. 4, no. 4, pp. 473–479, 1996.

[10] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *CVPR*, vol. 1, 2001, pp. 511–518.

[11] ——, "Robust real-time face detection," *IJCV*, vol. 57, no. 2, pp. 137–154, 2004.

[12] S. Sivaraman and M. Trivedi, "A general active-learning framework for on-road vehicle recognition and tracking," *IEEE Trans. Intell. Trans. Sys.*, vol. 11, no. 2, pp. 267–276, 2010.

[13] C. Caraffi, T. Vojir, J. Trefny, J. Sochman, and J. Matas, "A system for real-time detection and tracking of vehicles from a single car-mounted camera," in *IEEE Conf. Intell. Trans. Sys.*, 2012, pp. 975–982.

[14] A. Yoneyama, C. Yeh, and C.-C. Kuo, "Moving cast shadow elimination for robust vehicle extraction based on 2d joint vehicle/shadow models," in *IEEE Conf. Adv. Vid. and Sig. Based Surveill.*, 2003, pp. 229–236.

[15] Z. Sun, R. Miller, G. Bebis, and D. DiMeo, "A real-time precrash vehicle detection system," in *IEEE Workshop Appl. Comp. Vis.*, 2002, pp. 171–176.

[16] A. Bensrhair, M. Bertozzi, A. Broggi, P. Miche, S. Mousset, and G. Toulminet, "A cooperative approach to vision-based vehicle detection," in *IEEE Conf. Intell. Trans. Sys.*, 2002, pp. 207–212.

[17] A. Khammari, F. Nashashibi, Y. Abramson, and C. Laurgeau, "Vehicle detection combining gradient analysis and adaboost classification," in *IEEE Conf. Intell. Trans. Sys.*, 2005, pp. 66–71.

[18] G. Song, K. Leeand, and J. Lee, "Vehicle detection by edge-based candidate generation and appearance-based classification," in *IEEE Intell. Veh. Sym.*, 2008, pp. 428–433.

[19] T. Kowsari, S. Beauchemin, and J. Cho, "Real-time vehicle detection and tracking using stereo vision and multi-view adaboost," in *IEEE Int. Conf. Intell. Trans. Sys.*, 2011, pp. 1255–1260.

[20] T. Kowsari, S. Beauchemin, M. Bauer, D. Laurendeau, and N. Teasdale, "Multi-depth cross-calibration of remote eye gaze trackers and stereoscopic scene systems (to appear)," in *IEEE Intell. Veh. Sym.*, 2014.

[21] K. Tagaki, H. Kawanaka, S. Bhuiyuan, and K. Oguri, "Estimation of a three-dimensional gaze point and the gaze target from road images," in *14th International Conference on Intelligent Transportation Systems*, 2011, pp. 526–531.

[22] G. Andersen, "Focused attention in three-dimensional space," *Perception and Psychophysics*, vol. 47, no. 2, pp. 112–120, 1990.

[23] Y. Freund and R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *Computational Learning Theory*. Springer, 1995, pp. 23–37.

[24] W. Chang and C. Cho, "Online boosting for vehicle detection," *IEEE Trans. Sys., Man, and Cyber. B: Cyber.*, vol. 40, no. 3, pp. 892–902, 2010.

[25] B. Southall, M. Bansal, and J. Eledath, "Real-time vehicle detection for highway driving," in *CVPR*, 2009, pp. 541–548.

[26] P. Bergmiller, M. Botsch, J. Speth, and U. Hofmann, "Vehicle rear detection in images with generalized radial basis-function classifiers," in *IEEE Intell. Veh. Sym.*, 2008, pp. 226–233.

[27] Z. Sun, G. Bebis, and R. Miller, "Monocular precrash vehicle detection: Features and classifiers," *IEEE Trans. Im. Proc.*, vol. 15, no. 7, pp. 2019–2034, 2006.

[28] D. Alonso, L. Salgado, and M. Nieto, "Robust vehicle detection through multidimensional classification for on-board video-based systems," in *IEEE Int. Conf. on Im. Proc.*, vol. 4, 2007, pp. 321–324.

[29] H. Cheng, N. Zheng, C. Sun, and H. van de Wetering, "Vanishing point and gabor feature-based multi-resolution on-road vehicle detection," *Lecture Notes in Computer Science*, vol. 3973, pp. 46–51, 2006.