



Computer vision and driver distraction: Developing a behaviour-flagging protocol for naturalistic driving data

Jonny Kuo^{a,*}, Sjaan Koppel^a, Judith L. Charlton^a, Christina M. Rudin-Brown^b

^a Monash University Accident Research Centre (MUARC), Monash University, Australia

^b Human Factors North, Inc., 174 Spadina Avenue, Suite 202, Toronto, Ontario, Canada

ARTICLE INFO

Article history:

Received 1 April 2014

Received in revised form 2 June 2014

Accepted 6 June 2014

Available online 23 July 2014

Keywords:

Naturalistic driving
Driver distraction
Observational study
Computer vision
Video processing
Machine learning

ABSTRACT

Naturalistic driving studies (NDS) allow researchers to discreetly observe everyday, real-world driving to better understand the risk factors that contribute to hazardous situations. In particular, NDS designs provide high ecological validity in the study of driver distraction. With increasing dataset sizes, current best practice of manually reviewing videos to classify the occurrence of driving behaviours, including those that are indicative of distraction, is becoming increasingly impractical. Current statistical solutions underutilise available data and create further epistemic problems. Similarly, technical solutions such as eye-tracking often require dedicated hardware that is not readily accessible or feasible to use. A computer vision solution based on open-source software was developed and tested to improve the accuracy and speed of processing NDS video data for the purpose of quantifying the occurrence of driver distraction. Using classifier cascades, manually-reviewed video data from a previously published NDS was reanalysed and used as a benchmark of current best practice for performance comparison. Two software coding systems were developed – one based on hierarchical clustering (HC), and one based on gender differences (MF). Compared to manual video coding, HC achieved 86 percent concordance, 55 percent reduction in processing time, and classified an additional 69 percent of target behaviour not previously identified through manual review. MF achieved 67 percent concordance, a 75 percent reduction in processing time, and classified an additional 35 percent of target behaviour not identified through manual review. The findings highlight the improvements in processing speed and correctly classifying target behaviours achievable through the use of custom developed computer vision solutions. Suggestions for improved system performance and wider implementation are discussed.

© 2014 Published by Elsevier Ltd.

1. Introduction

Driving has been described as a ‘satisficing’ task, for which drivers will develop and invest only the minimum level of skill and attention required to complete it (Hancock et al., 2008). As such, driver engagement in secondary tasks has been found to be highly prevalent, with insufficient attention being directed at tasks necessary for safe driving. (Young and Lenné, 2010). Research has shown that up to 23 percent of crashes and near-crashes can be attributed to driver distraction, and that when drivers direct their gaze away from the forward traffic scene for more than 2 s, their crash risk is more than doubled (Klauer et al., 2006).

Extended periods of data collection through the use of discreet, in-car video cameras has allowed for naturalistic driving studies (NDS) to objectively capture aspects of everyday driving, including those that may be indicative of driver distraction, that were previously inaccessible to researchers (Klauer et al., 2006; Hanowski et al., 2005; Stutts et al., 2005). However, with the significantly increased volume of data generated comes the potentially challenging and inherently error-prone task of observation and interpretation by human analysts. This gives rise to both logistical and inferential limitations. Firstly, the manual processing of NDS data by human analysts becomes more time and labour-intensive with growing dataset sizes. Previous pilot research by the authors has yielded 150 h of video footage (Koppel et al., 2011), with current efforts aiming for 700 h (Sun et al., 2012). To limit the total amount of data that analysts need to view, one approach has been the use of various statistical sampling methods to select a subset of data from the complete dataset to analyse (Stutts et al., 2005; Koppel et al., 2011). Such protocols can be easily implemented without the

* Corresponding author at: Monash Injury Research Institute, Building 70, Monash University, Victoria 3800, Australia. Tel.: +61 3 9905 1808.
E-mail address: jonny.kuo@monash.edu (J. Kuo).

need for specialised hardware or software. However, fundamental statistical assumptions are made regarding the representativeness of the selected subset, the veracity of which is difficult to assess.

Another approach to data reduction has been the use of video triggers such as vehicle performance data (i.e. only reviewing epochs of video data that are temporally correlated with sudden braking or swerving manoeuvres, as recorded by vehicle 'black box'-type devices) (Klauer et al., 2006). However, while these critical incident-triggered epochs offer valuable insight into the proportion of crashes/near-crashes attributable to driver distraction, the prevalence ratios of driver distraction may not be validly inferred without also considering instances where the occurrence of driver distraction did not result in a critical incident. In addition to these reasons for specifically measuring the occurrence of driver distraction, distraction-triggered data (as opposed to crash-triggered) may provide unique insight into the mechanisms that differentiate incidents of distraction which result in crashes and those that do not.

Eye-tracking technologies such as FACELAB are an example of a distraction-centred approach to data reduction, using driver glance location as a surrogate measure of where a driver is directing his or her attention (Taylor et al., 2013). The implementation of dedicated eye-tracking hardware for NDS data collection has allowed researchers to gain a high level of detail relating to visual distraction of drivers as it occurs in their natural environment (Liang et al., 2012; Ahlstrom et al., 2012). However, eye-tracking using currently available solutions remains an *a priori* venture, requiring forethought in research design. While these tools offer researchers a high degree of fidelity in what they measure, datasets collected without such applications in mind (or before the development of these tools) remain incompatible and must rely on conventional manual coding protocols. This represents a significant underutilisation of resources, both in the quantity of data left unexamined and in the need for manual coding when automated techniques exist.

The application of machine learning and computer vision solutions to NDS data offers a promising approach to resolving the issues described above, with many sophisticated applications based on driver face-tracking developed in the field of computer science (Bergasa et al., 2008; Rezaei and Klette, 2011). However, few of these applications have been tested extensively with large NDS datasets. There is a need to develop machine learning solutions that are resilient to the inherent 'noise' in naturalistic data (Young et al., 2008), and that not only accommodate the physical and technical limitations of existing NDS datasets, but that may also potentially be used to analyse future datasets collected without the use of such dedicated hardware.

To address the challenges posed by manual coding protocols, the aims of the present study were to develop a computer vision solution for classifying driver glance behaviour captured using video-recording during NDS. Additionally, using results derived through manual coding from the Children in Cars data set (Koppel et al., 2011) as a benchmark of current best practice, a second aim was to compare the accuracy and speed of processing achievable by a computer vision solution. It was hypothesised that manual coding and computer vision approaches would differ significantly, with computer vision processing correctly classifying a greater number of off-road glances whilst requiring less processing time.

2. Method

2.1. Computer vision algorithms

Custom software was developed using Python programming language (<http://www.python.org>) and an open-source computer

vision library, SimpleCV (<http://www.simplecv.org>). These tools were selected for their high level of abstraction, allowing for rapid software development. Specifically the find HaarFeatures module of SimpleCV was used for face detection. This module is based on the Viola and Jones (2001) framework for face detection. The technique makes use of a series of adjacent dark and light rectangular regions (i.e. classifier cascades) to identify whether target features are present in an image. For a classifier cascade to be able to recognise a specified target feature, it must first be 'trained' by being presented with examples of what is a correct instance (a positive image) of the feature and what is an incorrect instance (a negative image). This training process is computationally intensive and is typically an application-specific process, requiring many thousand, manually selected and cropped examples of positive and negative images (Lienhart et al., 2002). Compared to the labour-intensive process of manual review, the selection of these training images need only be performed once and may subsequently be used to classify any number of images (given the same target feature and environmental conditions). For face tracking applications, researchers have proposed the training and use of multiple classifiers to account for different head positions and lighting effects (Jones and Viola, 2003). Head position specificity of classifier cascades was exploited in the present study as a robust method to identify instances when drivers turned their heads away from the forward traffic scene.

Perhaps due to the limited range of participant faces available in the present dataset, preliminary analyses showed more performance variability between participants than among different lighting conditions, suggesting the need for multiple participant-specific classifiers.

To this end, two approaches were implemented. In the first approach, separate classifiers were developed for male and female drivers. The decision to discriminate on driver gender was based on visual observation of a gender difference in hair styles which was hypothesised to manifest as highly salient differences in light and dark regions, as per the underlying mechanisms of the Viola and Jones (2001) technique. In the second approach, a statistical-based method was used. Hierarchical cluster analysis was performed on averaged images of each driver to determine the minimum number of classifiers that would need to be trained. Full results of this analysis are presented in Section 3.1. In brief, three clusters were identified: two clusters of three male drivers each, and one cluster consisting of six female drivers plus one male driver.

2.2. Datasets

The dataset from which the test and training sets were drawn consisted of 621 discrete journeys (i.e. 165 h of vehicle travel). A summary of the data management protocol is presented in Fig. 1.

Participant characteristics, recruitment, and procedure used in obtaining the data set have been previously documented (Koppel et al., 2011; Charlton et al., 2010). In brief, 12 families were recruited from an existing Monash University Accident Research Centre (MUARC) database on the basis of regularly driving at least one child between the ages of 1 and 8 years who were typically seated in a child restraint system (CRS) in the backseat. Families were provided with a luxury model family sedan for a period of three weeks, during which they were instructed to drive as per their usual routines. The study vehicle was fitted with four discreet colour cameras set to automatically record driver and passenger in-vehicle behaviours. The following perspectives were recorded through the video system: the forward traffic scene, a view of the driver and front seat passenger, the rear left passenger, and the rear right passenger.

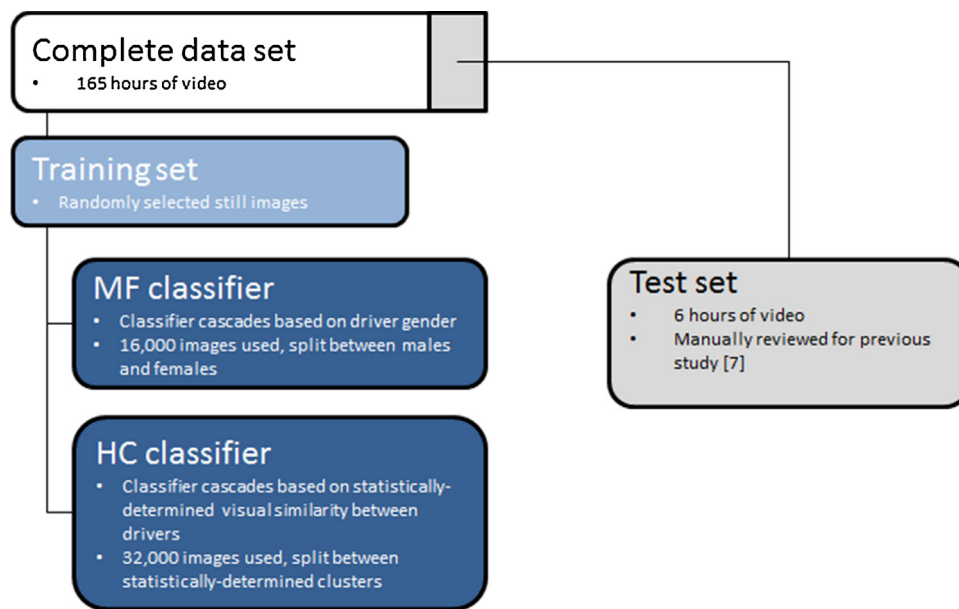


Fig. 1. Overview of data management protocol in the creation of the test and training sets.

2.2.1. Test set

The test set consisted of videos of 20 separate trips from 7 (3 male) drivers, totalling 6 h and 15 min of footage. The videos were selected based on the availability of corresponding second-by-second annotated Snapper (a video analysis and annotation tool, <http://www.webbsoft.biz/prod.snapper.php>) files used in our previous research, detailing precise information on behaviour typography, occurrence, and duration of driver distraction. For this data set, the criterion used to classify driver distraction events was any behaviour not directly related to the driving task (Koppel et al., 2011; Charlton et al., 2010). This included secondary behaviours in which drivers did/did not divert their eyes off road. For the purposes of the current study, behaviours where drivers' gaze remained on-road were excluded from the analyses. Instead, only behaviours involving the driver looking away from the forward traffic scene while the vehicle was in motion were selected for comparison. This included instances of the driver looking out the side window, turning around to check on rear seat passengers, as well as all instances of looking at the radio console.

2.2.2. Training set

Empirical analyses have suggested the use of 5000 positive images and 3000 negative images for optimal classifier cascade performance (Lienhart et al., 2002). Positive images denote images which depict the target object to be classified, in this context a driver's face directed away from the forward traffic scene. Conversely, negative images denote images depicting any other type of object or scene. For the male/female classifier set, 5000 positive images and 3000 negative images were selected for each gender. Similarly, for the hierarchical cluster analysis-based classifier set, 5000 positive and 3000 negative images were obtained for each cluster, equally distributed across the number of participants in each cluster.

To assist with the collection and cropping of training images, another classifier cascade was initially developed and trained using manually selected video frames from the test set. This cascade was developed as an expedient means to collect training images and thus was not comprehensively trained, with only a few hundred images used. Subsequently, for each driver, a video was randomly selected from the remaining data set. From these videos,

the expedient classifier randomly selected and saved positive and negative images. Due to the ad hoc training of this classifier (i.e. training was incomplete with only a few hundred images used), manual review of the saved images was required to ensure validity and reliability in creating the final training set.

2.3. Dependent measures

Classifier cascade performance was assessed on the degree to which classification of driver glancing behaviour was consistent with classification from manual review. Each cascade was compared with manual coding on the basis of the number of true positives, true negatives, false positives, and false negatives generated. Exemplars for each of these metrics are presented in Fig. 2. Additionally, the number of unique and correct classifications by the computer vision systems not otherwise classified in the test set by manual coding was also considered.

2.4. Hierarchical clustering of averaged driver faces

To determine the requisite number of classifier cascades to adequately account for all drivers, hierarchical clustering was conducted on averaged images of participant driver faces. Custom Python software was used to select and crop profile images of participant drivers from the video data (excluding videos from the test set). Using ImageMagick (<http://www.imagemagick.org>), sets of profile images for each participant were then averaged across the z-axis to create the averaged image, which was subsequently used in the hierarchical clustering analysis.

2.5. Procedure

Each video from the test set was analysed once with the hierarchical cluster-based classifier set and once with the gender-based classifier set. This returned a frame-by-frame output indicating whether the driver was directing their attention to the forward traffic scene. This output was then smoothed by averaging every 25 frames (based on the 25 fps frame rate) to generate a second-by-second output consistent with the resolution used in the Snapper project files. This output was then compared with the





Computer vision classification			
Manual coding classification	Eyes off road	Eyes off road	Eyes on road
		True positive	False negative
			
			* Participant is looking toward her lap
	Eyes on road	False positive	True negative
			

Fig. 2. Exemplar images of a true positive, false positive, false negative, and true negative in comparing manual review and computer vision software performance.

manually-coded Snapper project files for the relevant test set videos.

Processing of raw video footage by the classifiers was an automated process and was typically set up to run overnight or during off-peak hours, or on a standalone machine during business hours. Using an i5 2.70 GHz notebook computer with 8GB RAM, videos could be processed in real-time. As minimal human interaction was involved in initiating and maintaining the continued operation of the program, actual processing time of raw video by the classifiers was not included in processing speed comparisons. Rather, a comparison was made between reviewing the entire test set (as would be required under a manual coding protocol) and reviewing only instances of off-road glances flagged by the classifier (correct or not). In the absence of data describing the actual time expended by researchers in manually coding the original data set, the time required under a manual coding protocol was estimated to be at least equivalent to real-time replay of the test set.

3. Results

3.1. Hierarchical clustering

Pearson's *r* was calculated for each pair of averaged profiles to determine their visual similarity (i.e. the minimum number of classifier cascades that would need to be developed to correctly classify all drivers). Using 1-Pearson's correlation values and the Scipy scientific library for Python (<http://www.scipy.org>), hierarchical cluster analysis was conducted on these values and the following dendrogram plotted.

Based on the dendrogram in Fig. 3, a decision was made to form three clusters – C1 consisting of participants 9m, 10m, and 8m; C2 consisting of 9f, 8f, 4f, 2f, 6f, 4m, and 7f; and C3 consisting of 5m, 3m, and 7m. As participants 3f, 12f, 5f, and 11f were not represented in the test set, they were excluded from further analysis.

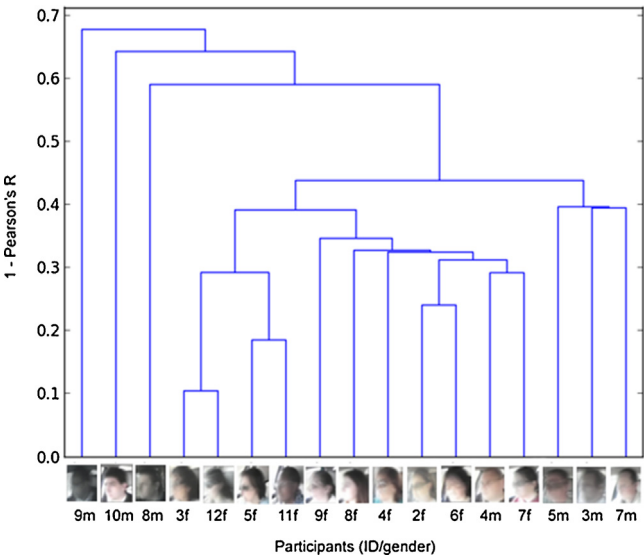


Fig. 3. Hierarchal cluster analysis of z-axis averaged profiles of participant drivers.

3.2. Processing time

A comparison of the total time required to complete the review process (manually-reviewed benchmark vs. the use of classifiers) is shown in Fig. 4. In sum, the test set consisted of over 6 h of video footage. False positives included, the use of the HC classifier set reduced the total duration of video to be manually reviewed by 55.84 percent, and the MF classifier set reduced it by 75.15 percent.

3.3. Classification accuracy

A comparison of the total number of off-road glances recorded among manual coding, the hierarchical clustering-based (HC) classifier set, and the gender-based (MF) classifier set is shown in Fig. 5.

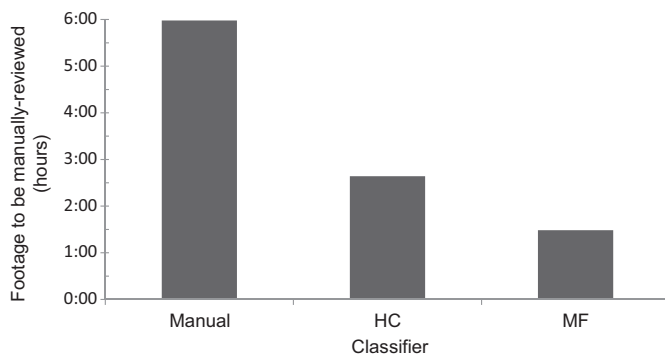


Fig. 4. Total hours of footage to be reviewed when manually coding, using a hierarchical clustering-based (HC) classifier set, and a gender-based (MF) classifier set.

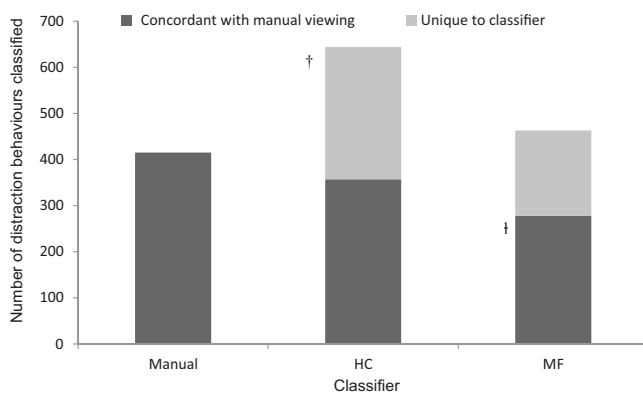


Fig. 5. Total number of off-road glances classified by manual viewing, hierarchical clustering-based classifiers (HC), and gender-based classifiers (MF). †Significant one-way repeated measures ANOVA between Manual, total HC, and total MF. ‡Significant one-way repeated measures ANOVA between Manual, concordant HC, and concordant MF.

HC and MF classifier sets achieved 86.02 percent and 66.98 percent concordance with the manual coding benchmark, respectively. One-way repeated measures ANOVA was conducted and showed statistically significant differences between the classifier sets' concordant hits and the benchmark, $F(1, 19) = 26.362, p = .000, \epsilon^2 = .581$. Pairwise comparisons of concordant hits showed that HC classifier output did not differ significantly from the results of manual coding, but MF classifier output differed significantly from both manual coding and HC, $p = .005$.

HC and MF classifier sets captured an additional 287 and 185 instances respectively of the target behaviour not otherwise captured through manual coding. These represented 69.16 percent and 35.18 percent of the total benchmark count. Repeated measures ANOVA showed a statistically significant difference between the benchmark and the total output from the classifiers, concordant and unique responses combined, $F(1, 19) = 36.932, p = .000, \epsilon^2 = .660$. Pairwise comparisons showed that the HC classifier differed significantly from both manual coding ($p = .005$) and the MF classifier ($p = .000$), but the MF classifier did not differ significantly from manual coding.

Additionally, not detailed in Fig. 5, HC and MF classifiers generated 157 and 146 counts, respectively, of false positives (i.e. flagged instances of off-road glances when none had actually occurred in video). Manual review of the misidentified instances revealed that the majority of these cases consisted of sporadic changes in lighting on the driver and atypical seating positions.

Comparisons were made between the individual cascade files which comprised the HC and MF classifiers (see Fig. 6). 'True positives' denote the proportion of classifications concordant with

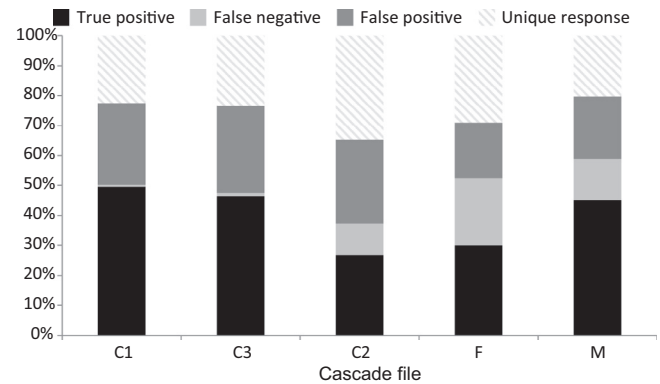


Fig. 6. Distribution of correct hits, misses, false positives, and unique responses within individual classifier cascade files.

manual coding, 'false negatives' represent occurrences of the target behaviour identified through manual coding but which were missed by the classifier cascade, 'false positives' denote epochs wrongly flagged by the classifier cascade as containing the target behaviour, and 'unique responses' mean correctly classified instances of the target behaviour that were not identified through manual coding.

Similarities in the distribution of correct classifications, false negatives, false positives and unique responses can be seen between C1 and C3, cascade files trained on 3 male participants each. Of particular note are the similar proportions of false positives and unique responses among all cascade files, with the major source of variation stemming from responses concordant with manual coding and false negatives.

4. Discussion

The aim of the present study was to compare the accuracy and overall processing time required of a computer vision classification solution versus manual coding in analysing naturalistic video data for classifying driver glance behaviour. The findings support the hypotheses that computer vision solutions can classify these events within NDS data at a faster speed than manual coding protocols while achieving comparable levels of accuracy. Furthermore, perhaps owing to the absence of observer drift or fatigue in a computer vision solution, additional instances of the target behaviour not otherwise classified by manual coding were identified.

A custom programed classifier set based on hierarchical clustering (HC) reduced the total duration of footage to be reviewed by 55 percent, whilst a gender-based (MF) classifier set achieved a 75 percent reduction in manual reviewing time. However, despite its advantages in processing time, output from the MF classifier differed significantly from the manual coding benchmark, achieving only 67 percent concordance. Additionally, the unique classifications made by the MF classifier not otherwise found through manual coding did not represent a significant increase from benchmark. In contrast, the HC classifier correctly classified 69 percent more instances of off-road glances beyond that captured by manual coding ($p = .000, \epsilon^2 = .660$), whilst also achieving 86 percent concordance. These results show strong support for the use of hierarchical clustering-based training for developing computer vision solutions to NDS data analysis.

With regard to classifier performance, the issue of false positives deserves particular mention. While the proportion of total classifications attributed to false positives may seem disproportionately high compared to the total number of classifications (circa 20 percent), the total amount of video footage to be subsequently manually reviewed, even with the inclusion of false positives, was

still considerably less than would be required by a manual coding protocol. HC and MF classifiers flagged 2.5 and 1.5 h of footage to be reviewed, respectively, whilst the test set contained 6 h of video. In reality, manual coding of the test set would likely have required significantly longer than 6 h, allowing for rewinding and replaying key epochs. Given the pre-flagged nature of epochs derived from classifier output, this issue would likely be much less pronounced when assisted by classifiers. While the selection of images to be used in training the algorithms required the equivalent of approximately one week's-worth of manual review, the resulting classifier could potentially be used for the reduction of an infinite quantity of video data given the same camera positioning. In contrast, the quantity of data yielded from a manual review process would only be directly proportional to the time and labour invested. Further examination of the workload involved in training different numbers of classifier cascades (and the subsequent effects on system performance) could potentially assist in optimising the process for developing similar solutions in future studies.

Additionally, fundamental differences exist between the nature of manual reviewing as a protocol in itself in comparison to manual reviewing output from the classifiers – manual reviewing as a protocol involves long periods of sustained attention by the analyst with additional time spent pausing and replaying key epochs of video. In contrast, with additional programming of video handling, manually reviewing classifier output could potentially be reduced to an analyst watching a pre-selected epoch of video and making a single binary decision as to whether or not a single instance of the target behaviour has occurred. This would reduce the cognitive workload required for manual coding, potentially allowing analysts to process more footage in a given timeframe. This would also reduce the need to train analysts on supporting tasks not directly related to recognising target behaviours, such as how to operate video analysis software packages.

In line with past research (Jones and Viola, 2003), a large proportion of false positives could be attributed to inconsistent lighting and atypical seating posture (whilst the driver continued to appropriately maintain their gaze on the forward traffic scene) which may not have been accounted for in the training set. This highlighted the difficulty with which classifier cascades generalise to stimuli other than the very specific exemplars used in their training. Future research could investigate the costs and benefits of including more varied training instances (e.g. across different postures and lighting conditions) and the effects this would have on over-fitting and minimising false positives. Conversely, the use of a separate classifier cascade to initially crop regions of interest containing a driver face (prior to assessing gaze direction) would reduce the data space considerably, potentially leading to greater system accuracy.

One of the limitations in the present study was the convenience test set, which was selected on the basis of available Snapper project files. The test set was not representative of all participants and all lighting conditions, and thus the extent to which the present findings apply to other data sets remains unknown. Given greater computing resources, more robust cross-validation techniques (e.g. leave-one-out, *k*-folds, etc.) may be attempted in future research. Regardless, the specificity of the algorithms used is likely to limit the application of the developed classifier cascades to datasets that utilise similar camera positioning within the vehicle. Additionally, while classifier assistance identified more instances of the target behaviour than manual coding, whether or not these new statistics are truly objective measures of off-road glancing frequency remain unknown as it is entirely probable for occurrences to be undetected by both manual review and computer vision analysis. As a hypothetical example, perhaps a participant consistently looks away from the forward traffic scene without moving their head – this would likely confound the performance of a head direction-based classifier. If, in addition to this, analyst fatigue whilst viewing the

same footage further confounded the formation of a manual coding benchmark, there would be no reliable measure against which the accuracy of the computer vision solution could be determined. While the present findings represent a quantifiable improvement on current best practice, objective performance of the classifier cascades remain unknown. Logically, the only valid way to truly determine algorithm accuracy would be through the use of a randomised, double-blind control design where the total count of distraction behaviour (or any other target behaviour) is determined a priori. Lastly, it should be noted that visual distraction comprises only a subset of the complete taxonomy of driver distraction, with many secondary behaviours not affecting gaze direction to the same magnitude as the target behaviour in the present study (if at all). Of the instances when drivers looked away from the road, it was not always possible to deduce the motivating factor for this behaviour. This issue could be addressed in future research through more comprehensive video coverage, including views of the vehicle dashboard or radio console.

As behavioural safety science begins to join other disciplines in leveraging vast quantities of data to answer its research questions, the development of a general use computer vision solution for data analysis represents a significant step towards bridging the gap between manual analysis and automated machine learning. Unanalysed data ultimately equates to squandered resources and missed opportunities for furthering research knowledge. The present findings demonstrate that significant improvements could be made in the data analysis process for NDS without the need for specialised computing or sensor hardware. Effective countermeasures, whether technological or educational, cannot be developed in the absence of domain expertise – in turn, domain expertise cannot be acquired without reliable and accurate observation.

Acknowledgements

The project is supported by the Australian Research Council Linkage Grant Scheme (LP110200334) and is a multi-disciplinary international partnership between Monash University, Autoliv Development AB, Britax Childcare Pty Ltd, Chalmers University of Technology, General Motors-Holden, Pro Quip International, RACV, The Children's Hospital of Philadelphia Research Institute, Transport Accident Commission (TAC), University of Michigan Transportation Research Institute and VicRoads. Additionally, we acknowledge the invaluable contribution of Chelvi Kopinathan, Samantha Bailey and David Taranto in conducting initial data coding to form the annotated dataset from which the present study derives, and to Prof Tom Drummond for expert advice on the computer vision implementation and directions for future research. Special thanks to Suzanne Cross for guidance and support in working with the retrospective dataset.

References

- Ahlstrom, C., Victor, T., Wege, C., Steinmetz, E., 2012. Processing of eye/head-tracking data in large-scale naturalistic driving data sets. *IEEE Trans. Intell. Transport. Syst.* 13 (2), 12.
- Bergasa, L., Buenaposada, J., Nuevo, J., Jimenez, P., Baumela, L., 2008. Analysing driver's attention level using computer vision. In: *Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems*, Beijing, pp. 1149–1154.
- Charlton, J., Koppel, S., Kopinathan, C., Taranto, D., 2010. How do children really behave in restraint systems. In: *Proceedings of the 54th AAAM Annual Conference*.
- Hancock, P., Mouloua, M., Senders, J., 2008. *On the Philosophical Foundations of the Distracted Driver and Driving Distraction*. CRC Press, New York City.
- Hanowski, R.J., Perez, M.A., Dingus, T.A., 2005. Driver distraction in long-haul truck drivers. *Transport. Res. Part F: Traffic Psychol. Behav.* 8 (6), 441–458.

- Jones, M., Viola, P., 2003. *Fast Multi-view Face Detection*. Mitsubishi Electric Research Laboratories.
- Klauer, S., Dingus, T., Neale, V., Sudweeks, J., Ramsey, D., 2006. *The Impact of Driver Inattention On near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data*. Technical report.
- Koppel, S., Charlton, J., Kopinathan, C., Taranto, D., 2011. Are child occupants a significant source of driving distraction? *Accid. Anal. Prev.* 43 (3), 1236–1244.
- Liang, Y., Lee, J.D., Yekhshatyan, L., 2012. How dangerous is looking away from the road? Algorithms Predict Crash Risk From Glance Patterns in Naturalistic Driving. *Hum. Factors: J. Hum. Factors Ergon. Soc.* 54 (6), 1104–1116.
- Lienhart, R., Kuranov, A., Pisarevsky, V., 2002. *Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection*. Intel Labs, Microprocessor Research Lab Technical Report.
- Rezaei, M., Klette, R., 2011. Simultaneous analysis of driver behaviour and road condition for driver distraction detection. *Int. J. Image Data Fusion* 2 (3), 217–236.
- Stutts, J., Feaganes, J., Reinfurt, D., Rodgman, E., Hamlett, C., Gish, K., Staplin, L., 2005. Driver's exposure to distractions in their natural driving environment. *Accid. Anal. Prev.* 37 (6), 1093–1101.
- Sun, Y., Papin, C., Azorin-Peris, V., Kalawsky, R., Greenwald, S., Hu, S., 2012. Use of ambient light in remote photoplethysmographic systems: comparison between a high-performance camera and a low-cost webcam. *J. Biomed. Opt.* 17 (3), 037005.
- Taylor, T., Pradhan, A.K., Divekar, G., Romoser, M., Muttart, J., Gomez, R., Pollatsek, A., Fisher, D.L., 2013. The view from the road: the contribution of on-road glance-monitoring technologies to understanding driver behavior. *Accid. Anal. Prev.* 58, 175–186.
- Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the Accepted Conference on Computer Vision and Pattern Recognition*, pp. 1–9.
- Young, K., Lenné, M., 2010. Driver engagement in distracting activities and the strategies used to minimise risk. *Saf. Sci.* 48 (3), 326–332.
- Young, K., Regan, M., Lee, J., 2008. *Measuring the Effects of Driver Distraction: Direct Driving Performance Methods and Measures*. CRC Press, New York City.