Road Scene Monotony Detection in a Fatigue Management Driver Assistance System

Luke Fletcher
Dept. Information Engineering, RSISE
Australian National University
Canberra, Australia
Luke.Fletcher@anu.edu.au

Lars Petersson
Autonomous Systems and Sensor Technologies
National ICT Australia
Canberra, Australia
Lars.Petersson@nicta.com.au

Alexander Zelinsky
CSIRO ICT Centre
CSIRO
Sydney, Australia
Alex.Zelinsky@csiro.au

***Abstract*— Automated fatigue detection devices show much promise in combating fatigue related accidents. One aspect which hampers the introduction of these technologies is context awareness. In this paper we develop and evaluate a road scene monotony detector. The detector can be used to give context awareness to fatigue detection tools to minimise false positives. The approach could also be used by road makers to quantify monotony on fatigue prone stretches of road. The detector uses MPEG compression to measure the change in information content of the road scene over time. We show that the detector correlates highly with human identified monotonous scenes. The technique is consistent over time and applicable for day and night operation. The compression is augmented with lane tracking data to distinguish between otherwise difficult cases. The detector is integrated into a Fatigue Management Driver Assistance System.**

I. Introduction

A great irony of transport systems research is that advances in road and vehicle safety can end up causing new threats to road users.
Drivers now enjoy:

- suspension design to minimise skeletal repetitive strain injury.
- sound damping to reduce road and traffic noise.
- mirror placement and instrumentation design to minimise driver effort.
- climate control to maintain a constant temperature regardless of the weather.
- cruise control to reduce the strain of driving over long periods.
- smooth low-curvature divided roads.
- multiple lanes or overtaking zones to reduce instances where drivers are stuck behind slower vehicles or need to overtake using the oncoming traffic lane.

In effect car manufacturers and infrastructure authorities have collaborated to attenuate stimulation from the off-driving tasks and ease the on-driving task. The unfortunate consequence is that drivers, now more than ever, are disengaged with the road environment other than the lane keeping task. If the task of lane keeping is under-stimulating, even for periods less than 20 minutes, the driver is susceptible to fatigue [1]. Consequently, sections of road that were once prone, for example, to head on collisions, are become fatigue accident zones (after divided multi-lane road redesigns). Ingersen et. al. [2] found that in Australia most fatigue accidents occur on a few high quality routes.

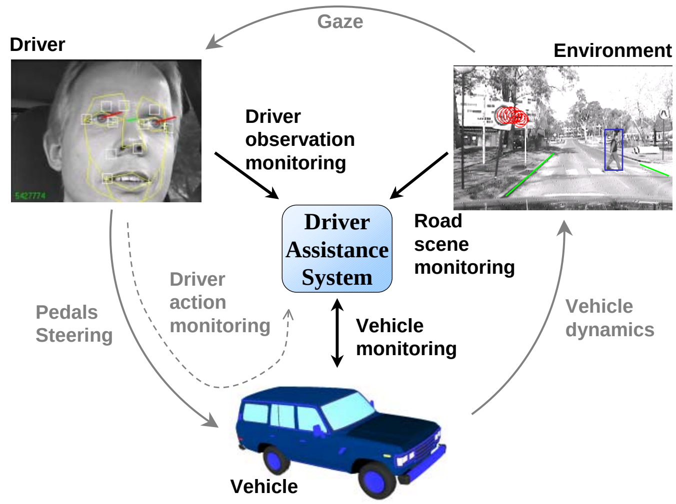


Fig. 1. Introducing driver observation to supplement driver action monitoring.

To address this problem, research has begun into how to monitor and maintain driver vigilance. Driver Fatigue detection and intervention is an active topic in the driver assistance systems. Unlike many other crash causes fatigue has proved notoriously hard to police. Driver log books are the primary tool used to manage fatigue in professional drivers however the systematic under reporting of driving hours remains common. In-vehicle driver monitoring has shown much promise for the detection of inattention and fatigue. Per-close eye closure, percentage road center eye gaze[3] and variance in steering wheel movement (SWM) are a few examples of metrics found to significantly benefit fatigue detection. One aspect which hampers these metrics is lack of context awareness. For example, many fatigue monitoring techniques struggle in urban and suburban driving scenarios; frequent blind-spot checking and intersection negotiation disrupt eye monitoring, frequent manoeuvring disrupts steering metrics. By detecting monotony in the road scene, fatigue monitors can be made context aware and thereby able to bias their metrics to minimise false positive

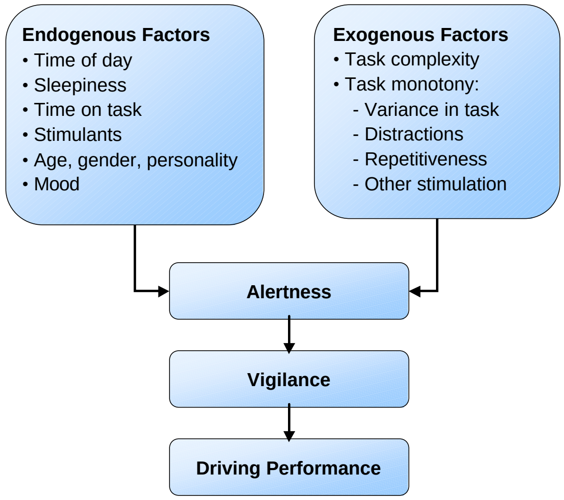


Fig. 2. Endogenous and exogenous factors contributing to fatigue [4].

warnings.

II. Driver Assistance Systems

Cars offer unique challenges in human-machine interaction. Vehicles are becoming, in effect, robotic systems that collaborate with the driver. As the automated systems become more capable, how best to manage the on-board human resources is an intriguing question. Combining the strengths of machines and humans, and mitigating their shortcomings is the goal of intelligent-vehicle research.

In the past many fatigue metrics use indirect driver monitoring. The driver's actions are measured and the driver's state was inferred. As vision systems have improved direct driver monitoring has become feasible. Per-close and Percentage Road Centre eye gaze metrics are the fruits of direct driver monitoring. Now we can go a step further, by integrating road scene data with driver data we can infer far more about the driver's behaviour - not only through the driver's actions, but also by the driver's observations (as illustrated in Figure 1).

III. Detecting Fatigue

The contributing factors of fatigue can be divided into endogenous (internal origin) and exogenous (external origin) sources. Lack of sleep can be considered an endogenous factor while lying in a darkened room would be an exogenous factor. Figure 2 shows the decomposition of contributing factors of fatigue. A recent trend in the psychology literature is to define monotony as an exogenous factor as opposed to a mental state (which would be endogenous, similar to boredom)[4]. In this way monotony can be used as an attribute of a task in a particular context. That is a task can be explicitly labeled as monotonous or non-monotonous (stimulating). The key point is that the monotony of the task can be decoupled from the actual mental state of the person. So regardless of how a task effects a person, if there are infrequent (or periodic) stimulus, low cognitive demand and low variance of task, it can be termed monotonous. The task of driving on a straight country road with little scenery on a clear day can be described as monotonous regardless of whether the driver is becoming fatigued or not. Whether a driver actually finds the trip fatiguing is dependent on the combined effect of the internal and external factors. A person driving home after being fired from their job is unlikely to become fatigued by a monotonous task.

Road Type	Scenery	Disruptions	Road Curvature	Monotony
Urban road	Cluttered	Frequent	High	**Low**
Country road	Moderate	Few	Varying	**Moderate**
Minor highway	Sparse	Varying	Moderate	**Moderate**
Major highway	Periodic	Varying	Low	**High**
Outback road	Sparse	Few	Low	**High**

TABLE I
Different driving environments and likely monotony level.

IV. Monotony Detection

As the primary sense used for driving, vision is also the primary sense to maintain alertness. We close our eyes and prefer a darkened room to sleep however sounds, smells and touch can be slept through. The monotony of the driving task can be decomposed into a visual stimulus component and non-visual component. As mentioned earlier the non-visual sensory stimuli have been attenuated by road and vehicle design, so we aim to measure the visual stimulus component.

Table I categorises a number of common road conditions into contributing factors for a monotonous environment. Figure 3 shows some sampled frames from a number of different road sequences. As experienced drivers, by observing the sampled frames a estimate of the monotony of the driving task for each sequence can be made.

To automatically measure the monotony in a road sequence we require a metric of the variance of the video sequence over time. In essence we need to estimate the information content of the sequence, the Kolmogorov-Chaitin complexity is just such an estimate. Kolmogorov-Chaitin complexity of a some original data is defined as the size of the smallest program (or encoding) which can reproduce the data [5]. That is the amount an information source can be compressed is a measure of the information content. Practically we need the measure that is robust against lighting variations common in outdoor scenes, periodic motion, and simple translational camera ego motion, MPEG encoding fills these requirements. MPEG encoding can be thought of as an estimate of the Kolmogorov-Chaitin complexity, though because the compression is lossy, it is not strictly valid. For our purposes the lossy nature of the compression and the effect on the metric is convenient as we will examine.

A. MPEG Compression as a measure of monotony

Moving Picture Experts Group (MPEG) encoding is a scheme for compressing a series of video or movie frames. MPEG exploits the property that in moving pictures only small

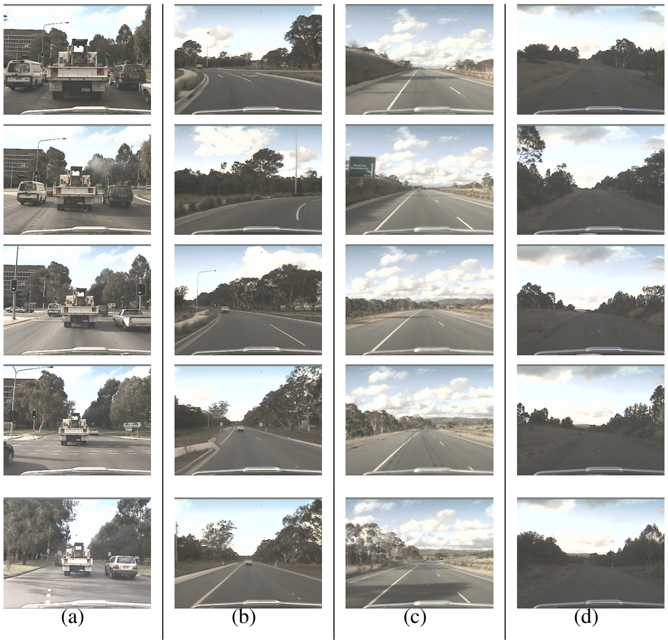


Fig. 3. Different driving environments with approximately 3 seconds between frames. (a): city traffic, not monotonous. (b): roundabout on otherwise straight road, not monotonous.(c): dual lane road gentle curvature, potentially monotonous. (d): no lane marks, not monotonous.

regions of the image actually change substantially between frames. Most of the frame is static or translates in the image, varying marginally. Impressive compression ratios are achieved by coupling an effective lossy compression scheme with an update procedure that can efficiently represent small motions and appearance changes between frames. Briefly, the video sequence is encoded as a sequence of three kinds of frames. Key (I) frames are compressed but otherwise complete frames using compression similar to but not identical to JPEG compression. Prediction (P) frames consist of the set of changes required to modify the previous frame into the current one. The frame consists of a set of motion vectors detailing how each subregion of the image has moved between frames and a correction representing what needs to be done to the previous subregion to make it match the current frame. Bidirectional prediction (B) frames are similar to P-frames except both previous and next frames are used to reconstruct the image. Benchmarking has shown that compression rates for each frame are typically approach: 1-bit/pixel for I-frames, 0.1 bits/pixel for P-frames and 0.015 bits/pixel for B-frames [6].

The compression and correction is done using the discrete cosine transform (DCT) which effectively is a frequency domain transformation. Motion between frames is measured using block matching techniques common to computer vision such as Sum of Absolute Difference (SAD) correlation. MPEG compressed sequences have been used in computer vision to:

- generate scene indexes of movies by the automatic detection of scene changes by large discontinuities between frames [7].
- detect pedestrians using P-frame motion vectors for optical flow [8].

MPEG4 compression has the following attributes which make it especially suitable for monotony detection:

- JPEG like compression of I-frames achieves better compression for "natural" images and worse compression for sharp edges, e.g. near-field cars compress badly.
- YUV color-space means some tolerance to lighting changes, a shadow across a region of the image compresses as scalar multiple of the subregion.
- Sum of Absolute Difference motion detection.
- With optimising encoders I-frames are included as required based on error margin. So dramatic changes in the scene like close moving vehicles will cause more I-frames to be added increasing the file size. One off scene changes, like entering a tunnel, will cause a single I-frame to be introduced.
- MPEG compression hardware makes embedded solutions easier.

MPEG compression as a measure of monotony is tempered by the following drawbacks:

- no guarantee that motion detection is capturing any meaningful motion in scene.
- fog or otherwise smoothed images compress well (though not monotonous).

B. Verification

To verify that MPEG encoding correlates with the monotony of a scene a monotony detector was implemented using the open source libavcodec library which provides an MPEG4 encoder.

Every 30th image was selected from the scene camera for compression. This represents a 1 second gap between frames. A sliding window of 150 images were compressed representing a time period of 2 minute 30 second window. The frames were 320x240 colour images and compression took around 1 second on a Pentium IV 3.0GHz machine. Compression was performed every 10 seconds. Most compression settings were left at the default values. Sum of Absolute Difference (SAD) correlation was selected for motion vector estimation. A high maximum bit rate of 8000 Kbit/s was selected allowing the compressor to use as much data as needed to encode high frequency changes in the image. Lower maximum bit rates forsake high frequency changes in the sequence to minimise the bit rate, which causes the kinds of changes we are interested in to be lost.

Table II shows the file sizes for various MPEG encoded sequences similar to those shown in figure 3. The encoded files have a good spread of sizes with a factor of two difference between the smallest and largest files. The MPEG/JPEG ratio shows that there is no correlation between the size of a JPEG sequence (representing only scene complexity) and the MPEG sequence (representing the change in the image over time).

When compared to a human judged monotony scale the MPEG file size has a strong correlation (see figure 4). The sole

Sequence	MPEG File Size (Kb)	JPEG Seq. Size (Kb)	MPEG/JPEG Ratio
Seq. G. figure 3(a)	3664	5168	0.71
Seq. B.	2976	4672	0.63
Seq. C.	2684	4460	0.60
Seq. N.	2604	10080	0.26
Seq. H. figure 3(b)	2548	4460	0.57
Seq. A.	2504	4324	0.57
Seq. E.	2412	4364	0.55
Seq. L.	2248	9836	0.23
Seq. D.	2176	4216	0.52
Seq. J.	2108	9352	0.23
Seq. I. figure 3(c)	2024	4452	0.45
Seq. M.	1972	9276	0.21
Seq. K. figure 3(d)	1784	8708	0.20

TABLE II

COMPRESSION OF VIDEO SEQUENCES.

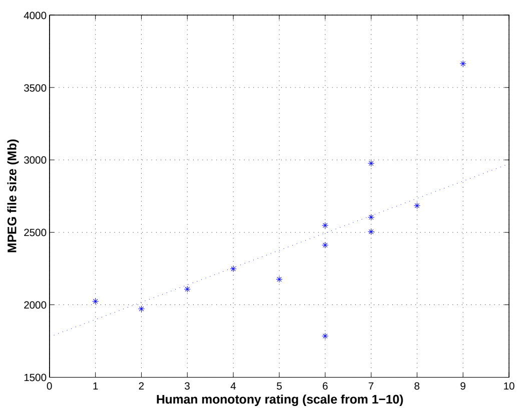


Fig. 4. Graph of various video file sizes versus a human evaluated monotony scale. 1 = very monotonous, 10 = very stimulating

outlier is the no lane markings sequence (figure 3(d)), which compresses very well but would have not been considered monotonous. The lack of sharp lane boundaries seems to allow a gentle transition between the frames. The course of the road is quite straight throughout the sequence (on par with figure 3(c)), but the lack of lane markings adds to the degree of difficulty to the lane keeping task decreasing the monotony.

C. Augmenting Compression with Lane Tracking

The primary failing of the MPEG monotony detector is in situations of poor visibility such as fog. The task is not monotonous yet the video will compress well. Detecting these cases would be possible as other groups have demonstrated systems capable of estimating the visibility of the road ahead. Hautiere et. al. [9] implemented a system that decomposed the luminance of the road ahead to judge the visibility range. As we have a previously developed lane tracking system we will use the lane tracking look-ahead distance as a similar measure.

The lane tracking subsystem is based on [10]. The system has been augmented to use a clothoidal road curvature model. A confidence measure is used to vary the look-ahead distance.

Fig. 5. Lane tracking look ahead distance varies with certainty.

When the variance of the primary road state variables (lateral offset, vehicle yaw and road width) increase beyond a small tolerance the look ahead distance is reduced to concentrate on robustly tracking the road directly in front of the vehicle at the expense of the far-field. As the near-field estimate converges the look ahead distance is increased. Figure 5 illustrates how road curvature estimate and look ahead vary.

The lane tracking look-ahead distance has the additional benefit in the monotony detection system of detecting other subtle cases such as crests (which may not show up as significant changes in the compressed sequence) and the gravel road case shown in figure 3(d). On a gravel road the lane tracker is still capable of tracking road using the colour difference between the road and the road shoulder and the weak edge information at the gravel boundary, but the increased uncertainty of the soft lane edges serves to keeps the look ahead distance low, indicating non-monotonous conditions.

V. FATIGUE COUNTERMEASURE DRIVER ASSISTANCE SYSTEM

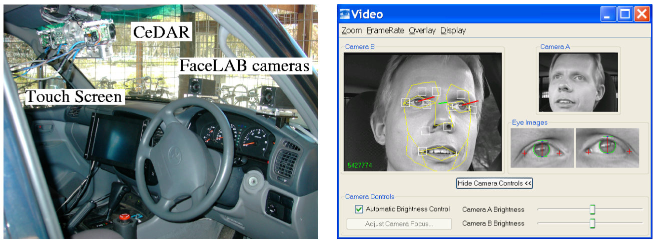
Per-close is a common metric often used to identify fatigue, it is the percentage of time the eye is closed over time window of several minutes. Fatigue is detected by checking for a critical maximum threshold of eye closure of the the time window. [3] uses the gaze direction to define the ”Percent Road Centre” distraction index. In this index the portion of time spent looking forward is measured over a time window. Unlike per-close this metric has an upper and lower bound for safe driving. If the index is too low the driver is likely to be distracted or falling asleep, if the index is too high the driver is staring not scanning the road scene indicating daydreaming or fatigue.

In this system we use Per-close and steering wheel movements (SWM) to detect fatigue. Fatigued steering motions are detected as a the mean steering wheel motion after subtration by a running mean. Mean steering wheel motions of more than 6 degrees over 5 minute intervals are declared fatigued.

A. Implementation

Previously [11], we introduced a driver assistance system framework designed around the key objectives of:

- Integrated video based driver monitoring.
- Video based road scene monitoring.
- Multi-hypothesis, ambiguity tolerant algorithms.
- Intuitive, unobtrusive, override-able, integrated driver assistance systems.



(a) (b)

Fig. 6. (a) The cameras in the test vehicle. The CeDAR active vision platform containing the scene camera and FaceLAB passive stereo cameras are labelled. (b) Driver head pose and eye gaze tracking using FaceLAB[13].

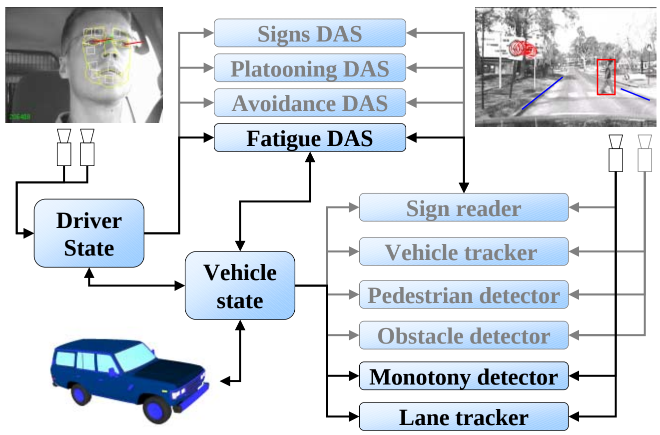


Fig. 7. Driver Assistance System architecture. Components used in this system are highlighted.

- Modular distributed architecture.

The framework has been used for integrating road scene understanding such as sign and pedestrian detection with the driver gaze for driver observation monitoring [12]. Figure 7 shows the system model for this application.

Figure 6(a) shows the scene cameras and driver monitoring cameras in the vehicle. The monotony detector and lane tracking system used a scene camera closest to the driver's head position. For this system the FaceLAB tracking system was used for the per-close estimate. FaceLAB is a driver monitoring system developed by SeeingMachines [13] in conjunction with ANU and Volvo Technological Development. It uses a passive stereo pair of cameras mounted on the dashboard to capture video images of the driver's head. These images are processed in real-time to determine the 3D pose of the person's face ($\pm$1mm, $\pm$1 deg) as well as the eye gaze direction ($\pm$3 deg), blink rates and eye closure. Figure 6(b) shows a screen-shot of this system measuring the driver's head pose and eye gaze.

The steering column is fitted with an encoder and strain gauges to measure the steering direction and torque.

B. Road trials

We conducted trials during the day, dusk and at night. Each time the vehicle was driven from the city out on an aterial

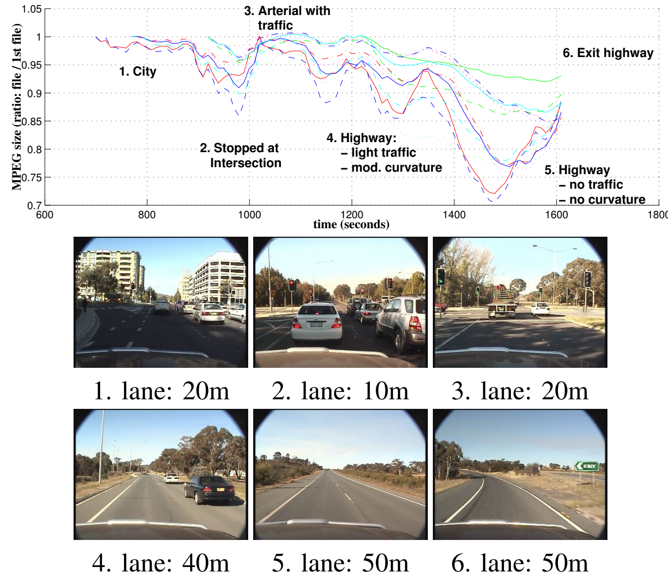


Fig. 8. MPEG compression and lane tracking look-ahead during afternoon trial on city and arterial roads. Sample images from the camera are shown at the corresponding numbered points with the lane tracking look-ahead distance.

road, onto a highway then back on a country road to the city.

To investigate how best to use MPEG encoding to represent monotony we encoded a set of movies every 20 seconds on varying the sampling rate and the length of the sequence. We trialed sampling at frequencies of: 4Hz (15[/60 frames]), 3Hz (20), 2Hz (30), 1Hz (60), 0.5Hz (120) with total durations of 10 seconds to 5 minutes. Figures 8 and 9 show the results of a day trial driving out from the city and back on the country road respectively. Figure 10 shows the result of a night trial along the same route. Overall the results were very promising. Both graphs show the largest trough in the MPEG file size when the car was stopped for a prelonged period at road works. Trends of smaller file size (or increased monotony) appear as the vehicle leaves the city for the highway and along the country road both during the day and at night. There is a good consistency across all MPEG frequencies and durations showing the monotony measure is well conditioned and not just an artifact of a particular sampling rate or total interval. As would be expected the smaller duration sequences are more responsive to brief changes in conditions while the longer sequences reflect the longer term trends. The faster frame rates seem to vary more regardless of the total durations indicating a better use of motion compensation. In the slower frame rates the vehicle motion between frames causes a significant difference in the image appearance which is to be too dramatic to capture using the motion compensation. The compression for these longer rates still represents a measure of the similarity of the scene over time, but more of the a general scene appearance instead of sequential motion. The lane tracking look-ahead distance was effective in identifying sections of road with a higher monotony level than expected by the MPEG compression alone. Particularly cases such as moderate

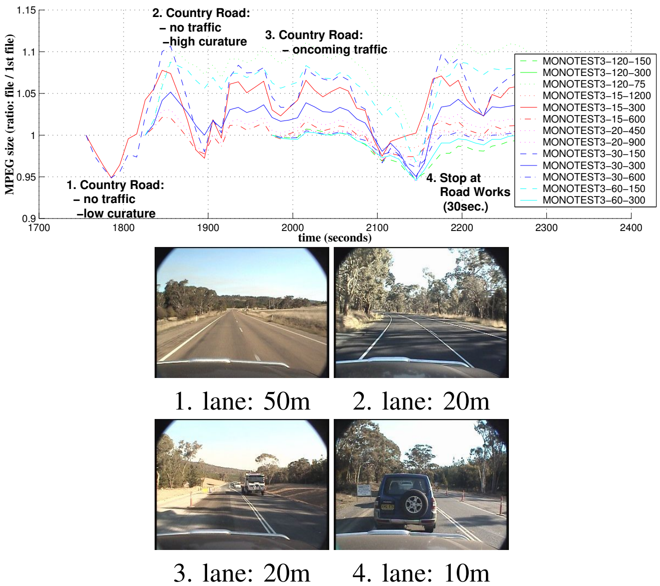


Fig. 9. MPEG compression and lane tracking look-ahead during afternoon trial on a country road back into the city. Sample images from the camera are shown at the corresponding numbered points with the lane tracking look-ahead distance.

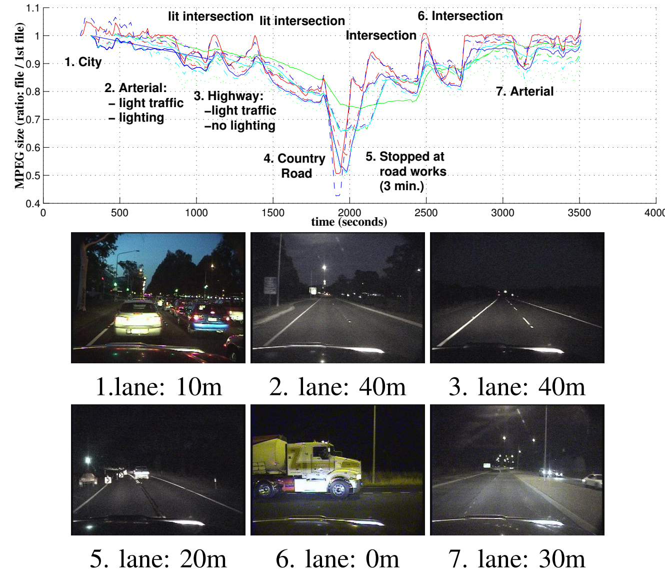


Fig. 10. MPEG compression and lane tracking look-ahead during a night trial on a city, arterial and country roads. Sample images from the camera are shown at the corresponding numbered points with the lane tracking look-ahead distance.

curvature country roads, crests and sections with no lane marks were identified as less monotonous than the compressed MPEG file would suggest.

The per-close and steering angle measurements showed produced a flat result showing the driver eye closer and steering angle variance was low indicating an alert driver which was correct and is the limit of such a small trial.

VI. Conclusion

A monotony detector has been proposed using MPEG compression to measure the change of information content of the road scene over time. The detector shows promise as a useful measure of the monotony of the driving task. Such a measure can be used to reduce false positives of fatigue monitoring systems. Though the correlation between MPEG compressibility and monotonous sequences is high there are some special cases such as low visibility weather and gravel roads that need to be handled explicitly. Using additional image analysis in the form of the lane tracking look ahead these cases can be managed. Finally, the monotony detector was integrated into a fatigue intervention driver assistance system and trials are under way.

A reasonably fast frame rate (2-4 Hz) over a long duration (¿5 minutes) seems to be the most promising for capturing monotony.

In future we hope to use detection to which reduce autonomous functions of the vehicle to raise driver alertness.

References

[1] P. Thiffault and J. Bergeron, "Monotony of road environment and driver fatigue: a simulator study," *Accident Analysis and Prevention*, vol. 35, pp. 381–391, 2003.

[2] I. P., "Tracing the problem of driver fatigue," *Driver Impairment, Driver Fatigue and Driving Simulation, Hartley, L. (Ed.), Taylor & Francis, London*, pp. 76–86, 1995.

[3] P. Larsson, "Automatic visual behaviour analysis," Master's thesis, Electrical Engineering, Linkping University, Sweden, Sweden, November 2002.

[4] P. Thiffault, "Environmental and personality precursors of driver fatigue and hypovigilance," November 2004, cARRS-Q 'Workshop on Monitoring Hypovigilance in monotonous conditions' presentation.

[5] G. J. Chaitin, "Information-theoretic limitations of formal systems," *J. ACM*, vol. 21, no. 3, pp. 403–424, 1974.

[6] Seeing Machines, "Mpeg information & frequently asked question page," *http:www.mpeg.orgMPEGstarting-points.html#faqs*, 2004.

[7] I. K. Sethi and N. V. Patel, "Statistical approach to scene change detection," in *Storage and Retrieval for Image and Video Databases (SPIE)*, 1995, pp. 329–338. [Online]. Available: citeseer.ist.psu.edu/sethi95statistical.html

[8] M. Coimbra and M. Davies, "A new pedestrian detection system using mpeg-2 compressed domain information," in *Proc. of IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP 2003)*, Madrid, Spain, 2003, pp. 598–602.

[9] N. Hautire and D. Aubert, "Driving assistance : Automatic fog detection and measure of the visibility distance," in *IEEE Intelligent Transport Systems*, Madrid, Spain, November 2003.

[10] N. Apostoloff and A. Zelinsky, "Vision in and out of vehicles: integrated driver and road scene monitoring," *International Journal of Robotics Research*, pp. 5–28, 2003.

[11] L. Fletcher, L. Petersson, and A. Zelinsky, "Driver assistance systems based on vision in and out of vehicles," in *IEEE Intelligent Vehicles Symposium (IV2003)*, Columbus, Ohio, USA, 2003.

[12] L. Fletcher, L. Petersson, N. Barnes, D. Austin, and A. Zelinsky, "A sign reading driver assistance system using eye gaze," in *IEEE International Conference on Robotics and Automation (ICRA05)*, Barcelona, Spain, April 2005.

[13] Seeing Machines, "Facelab face and eye tracking system," *http:www.seeingmachines.com*, 2004.