Driver Attention and Behavior Detection with Kinect

Cleshain Solomon and Zenghui Wang
Department of Electrical and Mining Engineering, University of South Africa, Florida 1710, South Africa
Email: {cleshain.solomon, wangzengh}@gmail.com

***Abstract*—Modern vehicles are designed to protect occupants in the event of a crash. However, passenger protection can be combined with collision avoidance. Statistics have shown that human error is the number one contributor to road accidents. Advanced driver behavior and monitoring systems have been developed by manufacturers in recent years and many have been proven to be effective systems in the prevention of accidents. However, these systems do not provide the complete solution and the systems only detect driver fatigue but do not offer gesture detection. In this paper, a system that detects driver fatigue and distraction has been developed using non-invasive machine vision concepts to monitor observable driver behavior. Moreover, cellular telephone detection was also considered. This paper also explores how driver error can be reduced through inexpensive technology (a Windows developer Toolkit Kinect) which observes driver behavior and reacts when certain unwanted patterns in behavior have been recognized. The experiments have been done to validate the efficiency of this new system.**

***Index Terms*—Kinect, driver fatigue, driver behavior, behavior detection**

I. Introduction

Research in computer vision started in the 1960s when the first PhD thesis in this topic was published [1]. Computer vision is in essence the science and engineering discipline concerned with making inferences about the external world from digital imagery. In simpler terms it is the ability of computer systems to interpret, process, analyze and understand the real world i.e. it's the ability of a computer system to see its environment. For example when we as humans examine a color image, we see all the different colors, shapes and textures. The brain is trained to give meaning and structure to the image. A computer sees raw data.

The key concept of computer vision is to give the computer system the capability to interpret an image and to make sense of it. Various forms of computer vision systems exist today with a variety of applications. Some of the applications are in biometrics (e.g. face recognition), human to computer interaction (e.g. Microsoft Kinect), robot vision systems (e.g. pick and place systems in manufacturing), image analysis (e.g. image reconstruction), vehicle safety (e.g. front and rear view cameras), security systems (e.g. airport security), and computer graphics.

Face detection technology is an application of computer vision and is concerned with detecting and localizing the shape and size of a human face but ignores everything else in the same environment. Face tracking is the process of continuous face detection and face recognition is the process of identifying a face by its unique features and characteristics. Eye tracking works on the same principle as face tracking. Eye tracking is the process of establishing where the eyes are directed at and it can additionally measure the characteristics of eye movement including the eye itself such as the pupil size [2].

Several Advanced Driver Assistance Systems with the feature of driver drowsiness detection has been investigated in the past and in recent years by various organizations, scholars and manufacturers. Volvo Cars filed a patent in 2003 for a method to detect and track a driver's head and eyes in relation to the vehicle internal space. The method quantifies the position of the driver's head against a reference point and tracks the orientation of the driver's head and eyes [3]. Volkswagen has a system called Fatigue Detection System which monitors eye blinking and estimates the probability of micro-sleep-episodes [4]. The University of Central Florida has developed a monitoring system, which they filed a patent for in 2002, based on digital cameras that monitor head motion and eye motion with computer vision algorithms for monitoring driver alertness and vigilance for drivers of vehicles and trucks [5].

The objective of the research project was to design an intelligent system that monitors driver behavior and attention in real-time with real-time response. The system monitors if a driver is focused on the road ahead. The secondary objective is to detect whether a driver puts a cellular telephone to his/her ear while driving. If the system detects irregular driver behavior it will signal an acoustic and visual warning on the instrument panel. The system consists of four components, namely a video streaming and capture system, face detection, computer vision algorithm and a warning or alarm signal.

II. Kinect Based Driver Attention and Behaviour Detection System

The recommended solution is a system that combines driver fatigue detection with detecting driver cellular phone usage. On detecting unwanted behavior while

Manuscript received March 13, 2015; revised August 27, 2015.

doi: 10.18178/joig.3.2.84-89

driving, the system will activate a warning signal reminding the driver to stay alert and be aware. The digital camera system that will be used is the Microsoft Kinect digital camera system. The Kinect was designed for Microsoft as tool to interact and communicate with a computer system using gestures. A gesture can be described as a movement of part of the human body, e.g. the movement of the hand from left to right, Fig. 1. Because there is no CPU inside the sensor, the data processing is executed on the host device by the Kinect driver [6]. With the Kinect for Windows developer Toolkit, Microsoft has made it possible for the hobbyist or professional developer to detect simple postures, track distances, relative positions, or the angles between the joints and integrate this into software applications. It is possible to detect the "Hello" gesture for example if one hand is higher than the head joint and the *X* and *Y* coordinates are not too far from each other [7].

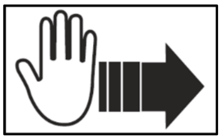


Figure 1. Hand gesture swipe from left to right.

The same principle is to be used in the design to determine whether the driver has a cellular telephone to the ear by determining the X and Y coordinates of the head joint and comparing that with the X and Y coordinates of the hand joint. If the hand and head joints are in close proximity for an unreasonable amount of time, the warning will be triggered reminding the driver to be attentive while driving. The system discards involuntary actions such as touching of the ear and face.

The head tracking algorithm is used to detect if the driver is nodding off or if the driver is about to fall asleep or has fallen asleep. The fatigue warning signal will be activated if driver fatigue has been detected. The warning will also be activated if the driver looks away from the direction of travel for an unreasonable time period.

III. Design of Driver Attention and Behaviour Detection System

In order to track the movement between two joints or more in a three dimensional space, the laws of trigonometry are used to calculate positions along the *X*, *Y* and *Z* axis's. The following needs to be determined:

1. Get the 3D coordinates
2. Create two Vectors and normalize them
3. Use the dot product method on the Vectors
4. Get the angles using Arccosine e.g. Math.Acos()
5. Determine the angle of rotation of the Vectors

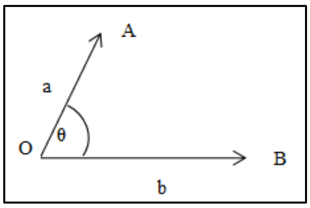


Figure 2. Vectors and representation of positions A and B in 3D space.

Fig. 2 shows the relative positions between point *A* and point *B*. With trigonometry the distance between these two points can be calculated. The angle between the points and the angle relative to the Kinect Sensor at point *O* can also be calculated.

To get the angle between the two points *A* and *B* (Fig. 2), (angle between the arm joint and shoulder joint), the scalar product (1) of a and b is used.

$$\boldsymbol{a} \cdot \boldsymbol{b} = |a| \times |b| \times \cos\theta \quad (1)$$

This is commonly known as the dot product for the two vectors *A* and *B* where θ is the angle between the vectors [8].

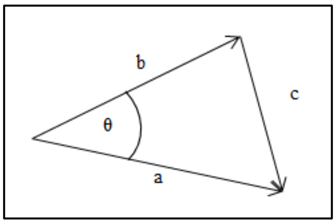


Figure 3. Vectors a, b, c.

By applying the law of cosines, and using the known vectors *a* and *b* seperated by angle θ (Fig. 3), the vector *c*, as in Fig. 3, is calculated (2) [8]:

$$c^2 = a^2 + b^2 - 2ab\cos\theta \quad (2)$$

In Fig. 4, the angle between two vectors is calculated by using the dot product of the vectors *A* and *B* (3) [9].

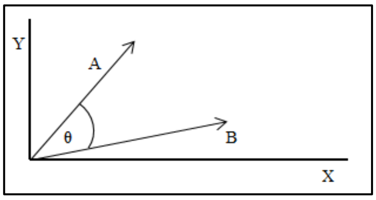


Figure 4. Dot product operator of two vectors.

From (1) we have:

$$\boldsymbol{A} \cdot \boldsymbol{B} = |A| \times |B| \times \cos\theta$$
$$\theta = \cos^{-1}(A \div |A| \cdot B \div |B|) \quad (3)$$

To generate a rotation in 3 dimensions, the axis of rotation and the amount of rotation have to be specified. With this the pitch angle ß, roll angle α and yaw angle *α* is determined. This can be used for tracking of the rotation of the head around the *X*, *Y* and *Z* axis. The yaw angle is the rotation of the head joint around the Y axis. The pitch angle is the rotation of the head joint around the X axis and the roll angle is the head joint rotation around the Z axis as shown in Fig. 5.

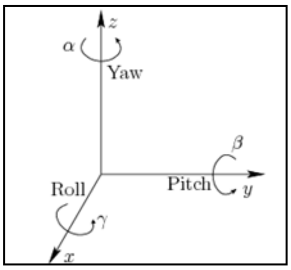


Figure 5. 3D rotation described as a sequence of yaw, pitch, and roll rotations.

Yaw angle is a counter clockwise rotation of α about the Z-axis. The rotation matrix related to yaw angle α is given as (4) [10]:

$$R_z(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (4)$$

In Fig. 6, Euler's rotation around *Z* is applied to determine the roll angle α.

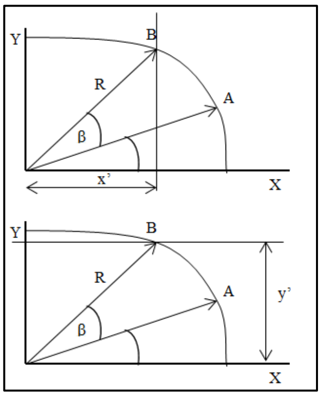


Figure 6. Euler's rotation around *Z*.

here:

$$x' = x\cos\beta - y\sin\beta \quad (5)$$

and:

$$y' = x\sin\beta + y\cos\beta \quad (6)$$

The pitch angle is the counter clockwise rotation of ß about the Y-axis. The rotation matrix related to pitch angle β is given as (7):

$$R_y(\beta) = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ \sin\alpha & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \quad (7)$$

Fig. 7 shows Euler's rotation around *Y* to determine the pitch angle ß.

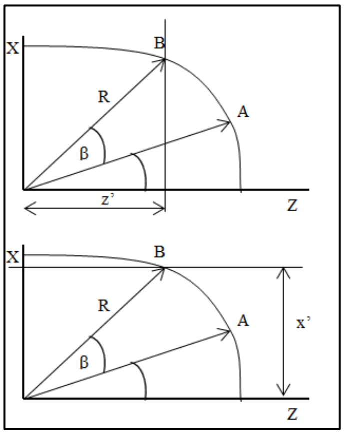


Figure 7. Euler's rotation around *Y*.

here:

$$x' = z\sin\beta + x\sin\beta \quad (8)$$

and:

$$z' = z\cos\beta - x\sin\beta \quad (9)$$

The roll angle is the counter clockwise rotation of γ about the X-axis. The rotation matrix related to roll angle γ is given as (10):

$$R_x(\gamma) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma & -\sin\gamma \\ 0 & 0 & \cos\gamma \end{pmatrix} \quad (10)$$

The following method, Fig. 8 applies Euler's rotation around *X* to determine the roll angle γ.

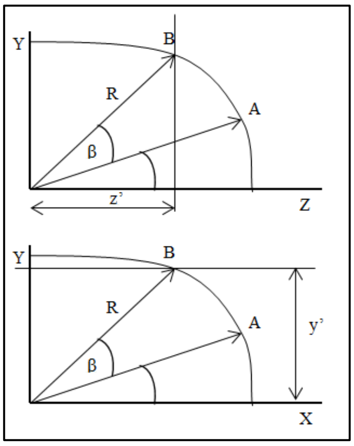


Figure 8. Euler's rotation around *X*.

here:

$$y' = y\cos\beta - z\sin\beta \quad (11)$$

and:

$$z' = y\sin\beta + z\cos\beta \quad (12)$$

Below, in Fig. 9 and in (13) are exaples of calculations, using some of the principles above, for Vectors *A* and *B* with angle θ. Here we calculate the distance between two joints. These are the distances between the head joint and hand joint or left shoulder joint. This procedure is for detecting whether the driver is using a cellular telephone while driving. The angle between the vectors can be determined (13).

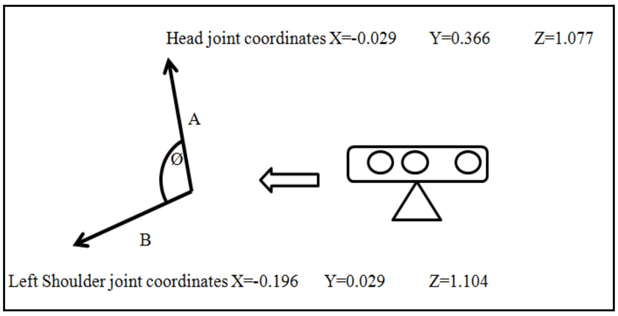


Figure 9. Illustration of projected vectors.

The vector and angle is calculated using the projected variables in Fig. 9 and Fig. 10 as following:

here:

$$\boldsymbol{A} \cdot \boldsymbol{B} = (A_x \times B_x) + (A_y \times B_y) + (A_z + B_z) \quad (13)$$

$$|A| = \sqrt{(0.029^2 + 0.366^2 + 1.077^2)}$$

$$\therefore |A| = 1.1378$$

$$|B| = \sqrt{0.196^2 + 0.079^2 + 1.104^2}$$

$$\therefore |B| = 1.1235$$

$$1.2236 = \sqrt{(1.2097)} \times \sqrt{(1.1235)} \times \cos\theta$$

$$\therefore \cos\theta = 0.99$$

$$and \quad \theta = 8°$$

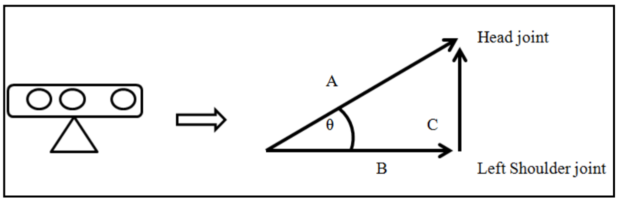


Figure 10. Determining distance between two joints.

And from (2) we have:

$$C^2 = A^2 + B^2 - 2AB \times \cos\theta$$
$$= 1.1378^2 + 1.1235^2 - 2(1.1378 \times 1.1235) \times \cos 8$$
$$C = 0.15838$$

The returned pitch, yaw and roll angles are used to determine whether the driver is looking away from the direction of travel. If the driver looks away for a predefined period, the warning is triggered to remind the driver to attend to the road ahead. The returned angles are also used to determine whether the driver is nodding off due to fatigue.

IV. Experimental Evaluation and Analysis

The first stage of the experimental evaluation was to test the Kinect Sensor performance and its feasibility for use as the digital camera system for fatigue detection. The initial test was to write and run the sample code which enables the colour stream, showing the video being captured, the depth stream, showing depth of the objects in front of the sensor and the skeleton stream that enables skeleton joint tracking, seen here in Fig. 11.

Next, was to enable skeleton tracking with joint tracking. The goal was to track the head joint and hand joints individually and return the 3D vector coordinates of the tracked joints. With the returned vectors the algorithm was designed to estimate the relative head pose and hand positions. The head pose is used to determine the head pitch, yaw and roll angles. These angles are used to determine whether the driver is nodding off or looking away from the direction of travel. The hand positions with respect to the driver's head position will determine whether a driver has a cellular telephone to his left ear.

Fig. 12 demonstrates how the joints are tracked. Here it can be seen that the tracked joints are used to calculate the vector distances between joints. For example in this demonstration the calculated distance between the left hand and the head joint on the *X* axis is 0.057m, on the *Y* axis the distance is 0.097m and on the *Z* axis the distance is 0.132m. The algorithm estimates that the driver has a cellular telephone in the left hand.

Figure 11. The depth stream and the colour stream.

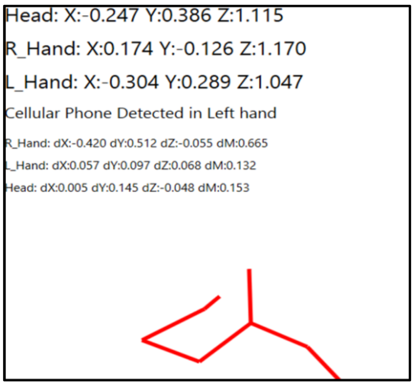


Figure 12. Demonstration.

The next stage of the development was to determine the head rotation and head pose angles. The driver's head pose is captured by three angles namely pitch, roll, and yaw angles. The driver's mouth position is also calculated. This determines whether the driver is yawning or not yawning. The following images, Fig. 13 to Fig. 20 show the returned data in real time. Here the pitch, yaw and roll angles can be seen. The returned value of the lower jaw is also shown. If the lower jaw travels down, the vector increases. When the returned vector goes beyond a specified value, the algorithm estimates that the driver is yawning and the yawn frequency is increased. If the yawn frequency goes beyond a predetermined range, the algorithm estimates that the driver is fatigued thereby triggering the first fatigue warning.

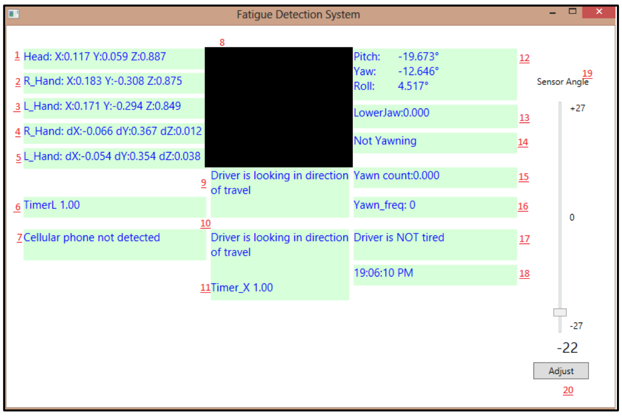


Figure 13. Fatigue detection system application window.

1. Head vector
2. Right hand vector
3. Left hand vector
4. Difference vector for right hand
5. Difference vector for left hand
6. Timer for left and right hand cellular phone detection
7. Cellular telephone warning message box
8. Video frame
9. Face direction indicator
10. Face direction warning
11. Face direction timer
12. Head rotation angles
13. Lower jaw travel distance
14. Yawning status indicator
15. Yawn counter
16. Yawn frequency
17. Driver fatigue warning message box
18. System time
19. Sensor angle tilt slider
20. Sensor adjust button

The next image in Fig. 14 shows the driver being tracked by the sensor. At this stage the driver is facing the direction of travel. The driver is not yawning and the yawn counter is zero. A cellular phone has not been detected. The sensor tilt angle is set at -3 degrees.

Figure 14. Driver not yawning with head facing direction of travel.

The following image in Fig. 15 shows that the driver has a cellular telephone in his hand. The cellular telephone timer is on 5 seconds at this stage. After 10 seconds, if the cellular telephone is still detected, the cellular telephone warning message will change to show that a cellular telephone has been detected. The timer is required to reduce false triggers and false warning messages.

Figure 15. Driver with cellular telephone in hand-no warning.

Fig. 16 shows that the cell phone has been detected with the warning message displayed. The warning message was triggered after 10 seconds. The timer resets to zero and will restart if the cellular telephone is still detected. The warning is repeated with an audible beep and a flashing red image. The flashing image and the beep will stop if the cellular telephone is no longer detected. Thereafter warning message will change to "Cellular phone not detected".

Fig. 17 shows that the cellular telephone has been detected in the other hand. The timer resets to zero and restarts.

The next image in Fig. 18 shows the head pose. Here it can be seen that the face direction information message has changed to "Driver is nodding up and down". The face direction timer is at 4 seconds. If this head pose stays at this detected pose for 10 seconds, the face direction warning message will change to show that the driver is not looking in the direction of travel. The same occurs if the driver looks to the left or right for a predetermined period of time.

Figure 16. Driver with cellular telephone in hand 1-warning displayed.

Figure 17. Driver with cellular telephone in hand 2-warning displayed.

Figure 18. Head pose down.

The following image in Fig. 19 shows that the face direction warning message has changed to "Driver is NOT looking in direction of travel". This occurs if the head pose is in this direction for more than 10 seconds. After 10 seconds the timer will reset and restart. The warning is repeated after another 10 seconds if the head pose does not change to direction of travel. The warning will stop if the head pose is in the direction of travel again.

Fig. 20 shows that the driver is yawning. A yawn will be detected if the lower jaw distance has travelled more than 0.05m. The yawn counter is at 29 yawns. The yawn frequency is at 3 yawns per minute. The driver fatigue warning message box has change to “Driver is Tired”.

Figure 19. Head pose warning message-driver looking down.

Figure 20. Driver yawning status.

V. Conclusion

A driver fatigue detection system which monitors observable behaviors was developed and it play an important role in monitoring driver actions and deciding whether the driver behaves in manner that is acceptable and safe. For distraction detection the head pose and or face direction was used to derive if the driver was looking on or away from the road ahead. For fatigue detection the drivers head pose angles were used. The detection and control system monitors pitch angle, roll angle and yaw angle. If any of the angles exceed a predefined value for a certain period of time, a warning is triggered, to remind the driver to be attentive. The angles were also used to determine if the driver was nodding off or whether he has fallen asleep. If a driver uses a cellular telephone while driving, the control system uses the *X*, *Y* and *Z* coordinates of the driver’s wrist joint and compares this coordinate and vector to the driver’s head position. The warning is triggered if the head position and wrist position are in close proximity of each other, but only after the positions do not change after 8 seconds. Experiments have showed the efficiency of the proposed system. The system excludes involuntary touching of the face, hair and ears and will consider this type of disturbances in the future.

Acknowledgment

This work was supported by China/South Africa Research Cooperation Program (No. 78673), South African National Research Foundation Incentive Grant (No. 81705).

References

[1] J. Aloimonos, “Purposive and qualitative active vision,” in *Proc. 10th International Conference on Pattern Recognition*, Atlantic City, NJ, 1990, pp. 346-360.

[2] A. Bojko, *Eye Tracking the User Experience*, New York: Louis Rosenfeld, 2013.

[3] P. Larsson and V. Trent, “Method and arrangement for interpreting a subjects head and eye activity,” USA Patent US7460940, 2008.

[4] T. V. Jan, T. Karnahl, J. Hilgenstock, and R. Zobel, “Don’t sleep and drive-VW’s fatigue detection technology,” in *Proc. 19th International Conference on Enhanced Safety of Vehicles*, Washington, DC, 2005.

[5] P. S. Smith, M. Shah, and N. D. V. Lobo, “Algorithm for monitoring head/eye motion for driver alertness with one camera,” in *Proc. International Conference on Pattern Recognition*, Barcelona, Spain, 2000, vol. 4, pp. 636-642.

[6] J. Abhijit, *Kinect for Windows SDK Programming Guide*, Mumbai: Packt Publishing, 2012.

[7] D. Catuhe, *Programming with the Kinect for Windows SDK*, 1st ed., Washington: Microsoft Press, 2012.

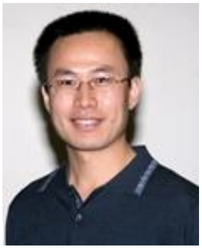
[8] K. Stroud, *Engineering Mathematics*, 5th ed., New York: Industrial Press Inc., 2001.

[9] R. Potter. A vector type for C#. [Online]. Available: http://www.codeproject.com/Articles/17425/A-Vector-Type-for-C#_rating

[10] S. LaValle, *Planning Algorithms*, Illinois: Cambridge University Press, 2006.

Cleshain T. Solomon has a Bachelor of Technology degree in Electrical Engineering at the University of South Africa, specializing in the field of Computer Science. He is a Technical Product Support Specialist at Volkswagen Group South Africa with more than 10 years’ experience. Previous work experience includes Instrumentation Technician at Volkswagen Group South Africa and Engineer-in-Training at the Coega Development Corporation.

Mr. Solomon is currently studying towards a Master of Technology degree and is a member of the South African Institute of Electrical/Electronic Engineers.



Zenghui Wang graduated with Doctoral degree (2007) in Control Theory and Control Engineering, Nankai University, China. He is currently a Professor in the Department of Electrical and Mining Engineering at University of South Africa. He has over 60 peer-reviewed journal and conference publications. His research interest is in the fields of Evolutionary optimization, Image and Video Processing, Adaptive control and Predictive control for nonlinear systems, Artificial Intelligence, and so on.