Application of Biologically Inspired Visual Information Processing in Affective Driver Status Monitoring

IL SONG HAN
WOO-SUP HAN
Korea Advanced Institute of Science and Technology
Korea, Republic of

Paper Number 15-0150

ABSTRACT

Recent times have seen an increased interest in technologies of driver assistance. Understanding the driver's current status is crucial for the implementation of Advanced Driver Assistance System (ADAS) and Driver Status Monitoring (DSM). Emotional factors such as anger have been long attributed to aggressive driving behaviours and increased likelihood of road accidents. Therefore, being able to accurately detect the affective states of the vehicle occupant will be critical for enhanced safety and comfort.

In this paper, we present a methodology for the evaluation of the emotional states of vehicle drivers. The proposed approach performs an assessment of the emotional states by using combination of biologically inspired visual information processing and neural networks coupled with feedback mechanisms. The system consists of the following stages: (1) biologically inspired image pre-processing; (2) facial feature extraction; (3) multilayer perceptron for classification; and (4) feedback mechanism. The system has been preliminary validated by using data available from Japanese Female Facial Expression (JAFFE) database. Four affective states were identified and tested, which includes anger, sadness, and happiness. Subsequent tests have shown the successful detection rate of 91.3% with test images, and over 70% correct classification in images with Gaussian noises, respectively.

INTRODUCTION

Recently there has been increasing attention towards automobile safety. According to the recent report by the UK government, road deaths have increased by 4% in the first quarter of 2014 compared to the previous year [**1**]. While the overall figure of fatalities has been under decline for the past decade, thanks to the improvement in medical treatment as well as the greater attention shown to the vulnerable road users such as the pedestrians and motor cyclists, the issue of road safety remains a key area for almost all nations in the world regardless of their wealth, with the fatality figure reaching over 1.2 million in the year 2010. Yet, only 28 nations in the entire world, which account for less than 450 million peoples of 7 billion world populations, are deemed to have sufficient legal measures to protect road users [**2**].

On the other hand, recent trends have shown that the developed nations are devoting a substantial amount of resources to improve road safety. In Britain, the government has spent £15.1 billion on the prevention of road accidents alone [**3**]. On a continental level, European Commission has launched a new EU Road Safety programme in 2010 as part of drive to cut the number of road fatalities by half between 2011-2020. All these findings show that the issue of road safety goes beyond the scope of industry.

Improving road safety comes in many different forms. The EU Road Safety programme has been divided into seven sub-sections: improving the education and training of road user; better enforcement of road rules; providing safer transport infrastructures; developing safer vehicles; promoting the use of modern technology to increase road safety; improving emergency and post-injuries services; and increased effort to protect vulnerable road users.

Within this paper, we have identified human emotions as one of the key influencers of driving performance. While most of existing Driver Monitoring System (DMS) focuses on drivers' concentration or drowsiness, another aspect of driver's status that deserves our attention is in emotion. Certain states of emotion, such as anger, and frustration can lead to aggressive and risky driving, which could result in accidents and fatalities. We believe that the accurate detection of emotional states from facial expressions is an important method to analyse and provide feedback to the vehicle driver's concentration level. Through successful recognition of emotional status, such as distraction, anger or frustration, it will be possible to improve the driving conditions and thereby safety of the vulnerable road users such as the pedestrians and cyclists.

We have applied image processing techniques and neural networks in the measurement of affective states based on still facial images in this study. The result of the finding can be used as a basis for future research in the application of affective computing in DMS.

There have been a number of studies which looked to present the feasibility of affective computing for improving road safety, and testing various methods of image processing in affective computing. It is notable that, despite numerous papers in the literature, there have been a minimal number of studies carried out on the use of image processing techniques coupled to neural networks in order to create an intelligent vision system which could evaluate human emotions.

METHODS

Biologically Inspired Orientation Filters

We claim that biologically inspired visual information processing offers a robust method for mimicking the robustness and flexibility of the primary visual cortex. One major strand of knowledge behind our current understanding of the behaviour of the primary visual cortex, an important brain area for vision, comes from the set of experiments by Torsten Hubel and David H. Wiesel [**4**]. Their experiment consisted of inserting microscopic electrodes into the visual cortex of experimental animals. This was used to read the activity of single cells in the visual cortex while presenting various stimuli to the animal's eyes.

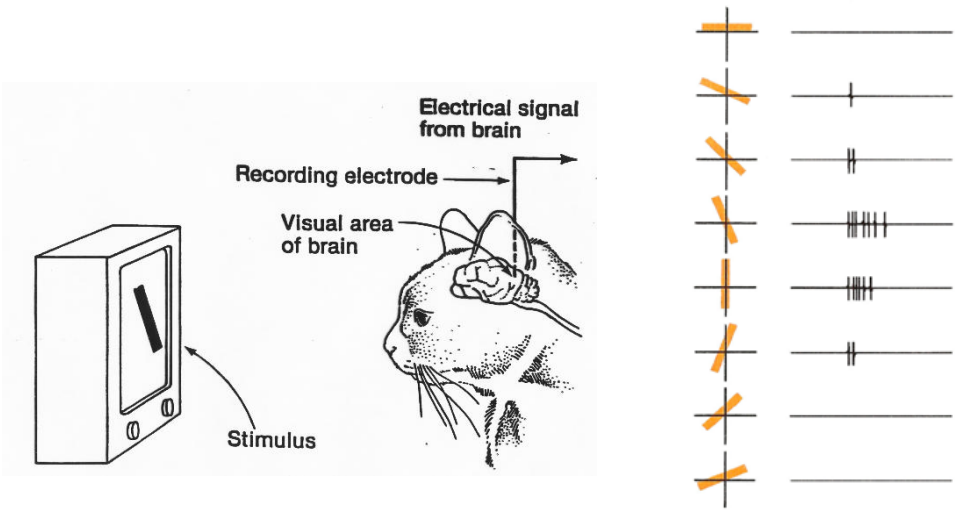

Figure 1. Hubel & Wiesel's Experiment Setup (left), and the responses of the cat's cortex when a rectangular slit of light of different orientations is shown (right)

They discovered that a topographical mapping in the cortex, i.e. that nearby cells in the cortex represented nearby regions in the visual field, i.e. that the visual cortex represents a spatial map of the visual field. Individual cells in the cortex, they found responded to the presence of edges in their region of the visual field. Furthermore, cells were found which would fire only in the presence of a vertical edge at a particular location in the visual field, while other nearby cells responded to edges of other orientations in that same region of the visual field. These orientation-sensitive cells were called "simple cells", and were found all over the primary visual cortex [**5**][**6**]

Based on Hubel and Wiesel's experimentation, the biologically inspired visual information processing incorporates the orientation selectivity of simple cell neurons to extract the features of facial images. The bahviour of the simple cells suggest that these cells possessed a patterned receptive field, with excitatory and inhibitory regions so that the cell would activate only if it received input (due to light) in the excitatory portion of its receptive field in the absence of input from the inhibitory portion. This operation is comparable to the operation of edge detection in image processing, which would process an image by spatial convolution with an edge filter [**7**].

We have simulated these biological operations through the application of multiple orientation filters for the feature processing of the facial expression. Our aim was to develop a filter which could accurately recreate the characteristics of the simple cells in mammalian visual cortex. These filters consist of six filters, each with specific orientation selectivity, yielding outputs of six orientation images for each input image (test image and reference image). Each set of six orientation filters provided orientation-selectivity for 0, 30, 45, 60, 90, and 135, degrees, and represents the receptive field properties of simple (linear) cells in V1.

Thus, the facial images were processed through this simulated simple cell function of the orientation selectivity and the six orientation images would then be combined to provide a resultant orientated image.

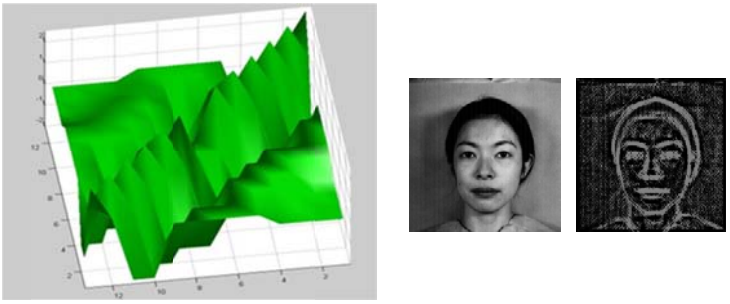

Figure 2. Image filter for 45 degrees orientation selectivity (left), and how the input image (middle) compares to the image processed by the image filter (right)

Affective Computing and Facial Action Coding System (FACS)

The resulting images have been then further processed to focus on the key 'regions of interest' (ROI). For determining the ROI, we have referenced results from contemporary researches in facial affect detection in the fields of affective computing. The detection and processing of facial expression has already been done through various methods such as, optical flow analysis, neural network, and active appearance model [**8**][**9**]. Facial affect detection can be done as a stand-alone affective computing system, and it can also be combined or fused alongside other modality (multimodal recognition, such as facial expression + speech pattern analysis or facial expression + hand gestures) to deliver a more robust analysis of the subject's emotional status. In this case, we have made use of Facial Action Coding System (FACS) for the purpose of determining ROI. FACS is one of the products of efforts to taxonomies human facial movements, and can be understood as an index list of facial expressions. Its foundation lies in Paul Ekman's cross-cultural research from the 1960s. His work suggested that affective expressions in face are not culturally determinant, but rather universal [**10**]. To support his points, Ekman carried out tests on Fore tribesmen in Papua New Guinea. The members of this tribe were not exposed to any contemporary emotional expressions, yet certain emotions – in particular six basic emotions: *anger*, *disgust*, *fear*, *happiness*, *sadness*, and *surprise*, were classified in the similar manner [**11**][**12**]. Thus, Ekman's study implied that certain facial expressions are biological in origin, and can be categorised with the same criteria.

While types of expressions which could be classified as the basic emotion have been under debate, the idea behind the various theories of emotion is the same. It suggests that the excluding core emotions, which could range between six to ten, all the other emotional states are simply synonyms of these core sets [**13**]. For our study, we have limited the emotive categories to be tested to four affective states of *happiness*, *sadness*, *anger* and *neutral*.

Thus, it is possible to demonstrate almost all facial expressions in terms of the fundamental actions of muscles, otherwise known as Action Units (AU) [**14**]. Every AU is independent of each other, and as such, a collection of AUs can be used to describe facial muscle movements including description of basic emotions. Ekman have incorporated FACS (Facial Action Coding System) into his research and provided corresponding AUs for each of the basic emotions, in addition to his own interpretation of their meaning. The six basic emotions of Ekman

would be described as shown in Table 1. In the context of affective computing, identification of emotional states using FACS code is known as EMFACS (Emotional facial Action Coding System), and they would concentrate on AU combinations related to emotional states. [**15**]

Table 1.
Basic Emotions and Corresponding Action Units

Emotion	**Action Units (AU)**
Happiness	6 ('cheek raiser') + 12 ('lip corner puller')
Sadness	1 ('inner brow raiser') + 4 ('brow lowerer') + 15 ('Lip Corner Depressor')
Surprise	1 ('inner brow raiser') + 2 ('outer brow raiser') + 5B ('Upper Lid Raiser*') + 26 ('Jaw Drop')
Fear	1 ('inner brow raiser') + 2 ('outer brow raiser') + 4 ('brow lowerer') + 5 ('Upper Lid Raiser') + 7 ('Lid Tightener')+ 20 ('Lip Stretcher') + 26 ('Jaw Drop')
Anger	4 ('brow lowerer') + 5 ('Upper Lid Raiser') + 7 ('Lid Tightener') + 23 ('Lip Tightener')
Disgust	9 ('Nose Wrinkler') + 15 ('Lip Corner Depressor') + 16 ('Lower Lip Depressor')

Determining Regions of Interest (ROI)

The actual process of calculating ROI has been done through detection of eyes and subsequent establishment of centre points between the eyes as the reference point through template based object recognition of eyes. From here, relevant AUs can be analysed for both the input image and neutral 'reference' image to which it would be compared against, as shown below. (**figure!**)

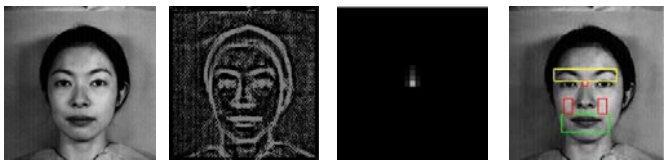

Figure 3. (from left) Input image, orientation image, detection of central point between two eyes as the reference location of calculating the ROIs, and how the calculated ROIs appear in the original input image.

Tests show that eye locations were detected successfully in both frontal and side facial image, demonstrating the effectiveness of combination of orientation selective feature extraction and feature based object recognition. This would, in turn, allow for measuring the level of driver attentiveness in road driving

situations. Subsequently, the detection of emotion (happiness, anger, or etc.) was based on the evaluation of processed facial features of selected areas.

Neutral Network

One of the requirements in developing a reliable warning system would be to provide feedback to the driver, whenever the driver would go into an extreme emotional state. To achieve this aim, it was important to detect which emotional state was the subject under, as certain emotive states are more counterproductive to safe driving than the other. However, more crucial aspect was in ensuring that neutral states of the subject was not being misinterpreted as an emotional state of any category. In this regards, an output of 1 (safe / neutral state) or 0 (unsafe / emotional state) would provide a sufficient result. The proposed system is designed to improve the successful detection of safe or unsafe emotional status of a driver, by tuning the ROI salience data encoding with the weight to the neural network input for improving the ordinary neural networks of Multi-Layer Perceptron (MLP).

Our input data would consist of analysed ROIs from the test and the reference images. This would be yielded by the data set of 90 images processed by the multiple orientation filters. The six data extracvted from six ROI were applied to the MLP to be classified into four emotional status, i.e. 'Neutral', 'Angry', 'Happy' and 'Sad'. The MLP of two hidden layers is designed to produce 4 outputs, which was implemented using the MATLAB neural network tool box.

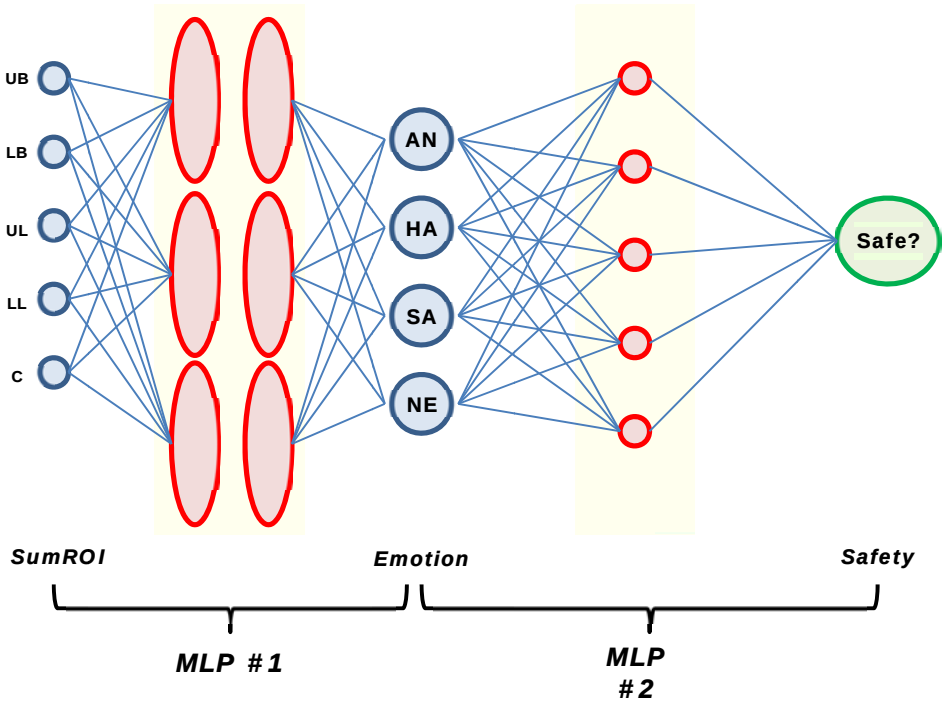

Figure 4: Design of proposed neural network system

During the design phase, the optimum network structure of 5x10x10x4 was determined by iterative trials of various network architectures with the observation of weight value distribution. The image processed by multiple orientation filters is assumed to be consistent in facial direction, size or scale.

The purpose of monitoring the driver's emotional status is attained by detecting the neutral emotional status as *safe* and the other status as *unsafe*, using the MLP of 4x10x5x1. The upper part of networks (4x10x5x1) is trained by an ideal situation, while lower part of networks (5x10x10x4) is trained by the back propagation learning using processed image data set.

Feedback Mechanism

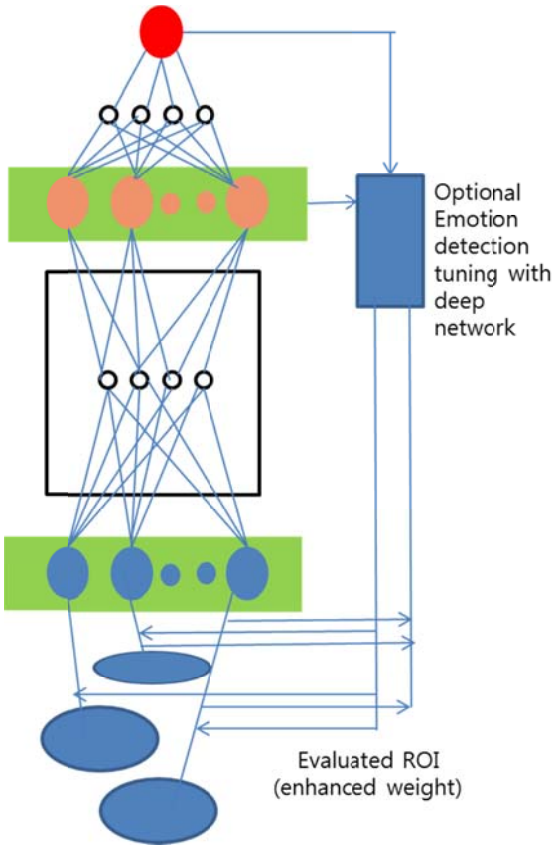

Figure 5. Proposed emotion classification system with weight tuning or learning at an early stage.

We stated in the introduction that our initial purpose was to develop a warning system which could provide an alert to the driver, whenever the driver would go into an extreme emotional state. To achieve this aim, it was important to detect which emotional state was the subject under, as certain emotive states are more counterproductive to safe driving than the other. However, more crucial aspect was in ensuring that neutral states of the subject was not being misinterpreted as an emotional state of any category. In this regards, an output of 1 (safe / neutral state) or 0 (unsafe / emotional state) would provide a sufficient result. The proposed system is designed to improve the successful detection of safe or unsafe emotional status of a driver, by tuning the ROI salience data encoding with the weight to the neural network input for improving the ordinary neural networks of MLP.

In addition to cascaded neural networks of MLP explained previously, the feedback to ROI is applied by the trained upper and lower networks of MLP. In case of failed detection of safety monitoring output, the encoding of ROI values to neural input is subject to the feedback with the lower weight, for the pre-analysed ROI for the largest value of emotional status – which is the main contributor to the detection failure. The feedback rate to ROI applied to neural network is a fixed value and the selection of ROI is controlled by the ROI value and the input status of upper network when feedback is required, for this framework. The method of deep network is scope of investigation by evaluating the effectiveness and relevance to the performance of the target function.

DATA SOURCES

The accuracy of the emotion detection was measured by tests carried out on the JAFFE database, through processing with two feedforward neural networks (multi-layer perceptrons) and the feedback mechanism based on a deep learning concept. The biologically inspired visual information processing showed a significantly high accuracy to emotion recognition without the need for precise matching or complex computation. This was superseded by mimicking the primary function of the simple cell of visual cortex, which provided a degree of robustness by maintaining accuracy even in test images with Gaussian noises.

RESULTS

Based on the data set of 360 images, the overall rate of detection returned over 91.3% accuracy, as shown in the Table 2 below. The proposed method of combining biologically inspired visual information processing with multilayer perceptron also gave a satisfactory performance, maintaining a 70+% accuracy level up to 5% Gaussian noise level. Investigation of deep-learning inspired feedback mechanism also yielded an improved performance, resulting in 27.2% reduction in the occurrence of false-positives.

Table 2.
Result of Biologically Inspired Visual Information Processing & Multilayer Perceptron

Input Class	Output Classification Result				Accurate Classification Rate
	Happy	Angry	Sadness	Neutral	
Happy	81	1	7	1	90.0 % (81 / 90)
Angry	0	90	0	0	100.0 % (90 / 90)
Sad	2	2	78	8	86.7 % (78 / 90)
Neutral	2	0	8	80	88.9 % (80 / 90)

Compared to the comparative studies carried out in affective computing, combination of biologically inspired visual information processing and MLP have outperformed methods based on artificial neural networks (ANNs; 84% accuracy) and combination of Sequential Floating Forward Search (SFFS), Fisher Projection (FP), and K-nearest neighbour algorithm (KNN; 81% accuracy) [**16**].

In depth review of biologically inspired visual information processing approach revealed certain patterns of interest. It was demonstrated that four emotional states (happiness, anger, sadness, and neutral) can be successfully classified with a small number of variables. Yet there were some level of differences as to how accurately each emotion was classified by biologically inspired visual information processing approach.

Angry emotional state was most likely to be detected with accuracy, with 100.0% successful detection rate, followed by happy emotional state with 90.0% classification accuracy. Sad emotional state had the most inaccurate incidents of classifications, which has also been documented in previous researches [**17**]. Reason for this difficulty could be attributed in parts to the comparative lack of facial expression in sad emotional state.

Compared to anger, which involves composite movements of eyebrows, competitive movements of eyelids, and major movement in lip areas, sadness would be expressed by movement of eyebrows and small movement to the corner of the lips. Even in real-life situations, detecting expression of subtle sadness is a challenge, and it is possible to infer such factors have been represented in the test results as well.

Further tests were carried out to determine the robustness of biologically inspired visual information processing approach, through introducing Gaussian noise to the images. This was done in order to examine its applicability in real life situations, where the image data is often corrupted. The results of classification with noisy test images showed a reduction of 16.01% in detection rate at 1% noise level, and maintained over 70% level of accuracy even at 10% noise level.

Table 3.
Detection rate (%) of biologically inspired visual information processing method with different level of Gaussian noise

Noise Level	SumROI(G)	SumROI(M)
0.0%	78.7	91.7
1.0%	76.4	79.1
2.5%	73.1	75.7
5.0%	68.8	73.5
10.0%	64.1	65.9

DISCUSSIONS AND LIMITATIONS

In this study, our proposed model for identifying affective states required the presence of a neutral emotive state, which would be used as a reference. While it is theoretically possible to process an individual's emotional status from a single image without any reference, in reality, there is a wide variability to the strength of human expression even within same emotive state. This variability is deemed dependent on the cultural, national, regional and gender difference of the subject [**18**]. This had also been reinforced during the analysis of datasets: that absolute values of expression within an individual cannot be used as a reliable measure. Further tests and statistical analysis could be carried out in the future as a means to develop an accurate and reliable indicator of emotive states based on input image alone.

Another challenge we have recognised during the experiments was that there are at least two kinds to the expression of emotions – genuine and fake. A fake expression often differs from a genuine expression. For example, in case of happiness, only the zygomatic major muscle, which runs from the cheekbone to the corner of lips, moves in case of fake emotional status. On the other hand, a genuine expression of happiness would involve movement of orbicularis oculi and pars lateralis (eyebrows) as well as zygomatic major. In addition to limited facial muscle movements, fake emotive states are also known to contain a certain degree of asymmetry [**19**],[**20**].

On the other hand, for certain emotional states in JAFFE dataset, there were very little expressions present. In comparison to the neutral image, even human vision had difficulty in determining the emotional state without the label.

Further investigation of other facial affection database has shown that this issue isn't unique to JAFFE database, as demonstrated in the figures below.

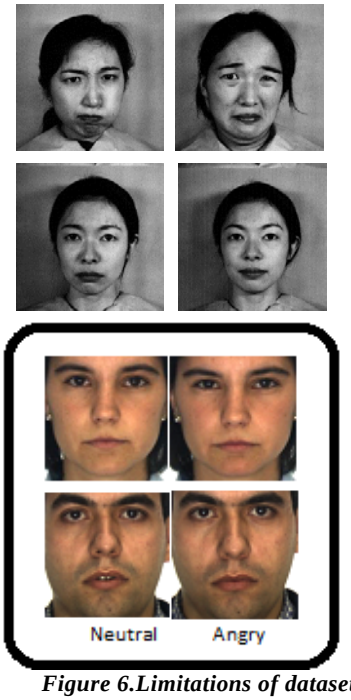

Figure 6.Limitations of dataset
Top Row: Examples of 'fake' emotion – asymmetry is clearly visible in the fake facial expression
Middle Row: Images displaying supposed expression of sadness (left) and neutral (right).
Notice the lack of difference between this image and that of the neutral state (right).
Black Box: Example of angry and neutral emotion from an alternative database.
Without the labels, it would be challenging to determine that left image is demonstrating anger and the right one is of the neutral emotional state.

CONCLUSIONS AND RELEVANCE TO THE SESSION SUBMITTED

We believe this research has produced a foundation from which further studies could be carried out. For example, would be in refining the biologically visual information processing approach. We have recognised that there are at least two kinds to the expression of emotions – genuine and fake. Within the context of driver monitoring system, fake emotions are unnecessary and should be ignored. Further studies carried out with images of emotions in real-life situations could provide a better training for the monitoring system capable of demarcating two categories.

Secondly, we have worked with the assumption of only neutral emotional states allowing for ideal driving situations. This hypothesis is under discussion within affective computing research, and will be given further updates in the coming time.

ACKNOWLEDGEMENT

The research was partly promoted as part of 'Technology Development for Behavior Improvement and Violation Control of High Risk Drivers' that was carried out by Ministry of Land, Infrastructure and Transport in Korea.

REFERENCES

[1] Reported Road Casualties in Great Britain: Quarterly Provisional Estimates Q1 2014. (2014). London: Department for Transport.

[2] World Health Organization. (2013). *WHO global status report on road safety 2013: supporting a decade of action*. World Health Organization.

[3] Reported Road Casualties in Great Britain: 2012 Annual Report. (2014). *Statistical Releases.* London: Department for Transport.

[4] Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *The Journal of Physiology*, 160(1), 106.

[5] Han, W. S., & Han, I. S. (2014). All Weather Human Detection Using Neuromorphic Visual Processing. In *Intelligent Systems for Science and Information* (pp. 25-44). Springer International Publishing.

[6] Han, I. S. and Han, W. S. (2014). Robust Detection of Pedestrian/Cyclist by Neuromorphic Visual Processing, In *FISITA 2014 World Automotive Congress, 2-6 June 2014.*

[7] Jhuang, H., Serre, T., Wolf, L., & Poggio, T. (2007, October). A biologically inspired system for action recognition, *IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007.* (pp. 1-8). IEEE.

[8] Picard, R. W. (2010). Affective computing: from laughter to IEEE, *IEEE Transactions on Affective Computing*, 1(1), 11-17.

[9] Hudlicka, E. (2003). To feel or not to feel: The role of affect in human–computer interaction. *International Journal of Human-Computer Studies*, 59(1), 1-32.

[10] Ekman, P. (1992). An argument for basic emotions, *Cognition and Emotion*, 6(3-4), 169-200.

[11] Ekman, P. E., & Davidson, R. J. (1994). *The Nature of Emotion: Fundamental Questions*, Oxford University Press.

[12] Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations, *Psychological Review*. 99(3):561-5

[13] Beck, J. S. (1995). *Cognitive Therapy*. John Wiley & Sons, Inc.

[14] Ekman, P., & Rosenberg, E. L. (eds.). (1997). *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS).* Oxford University Press.

[15] Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis, *Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000.* (pp. 46-53). IEEE.

[16] Kolodyazhniy, V., Kreibig, S. D., Gross, J. J., Roth, W. T., & Wilhelm, F. H. (2011). An affective computing approach to physiological emotion specificity: Toward subject-independent and stimulus-independent classification of film-induced emotions, *Psychophysiology*, 48(7), 908-922.

[17] Picard, R. W., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10), 1175-1191.

[18] Ekman, P., & Rosenberg, E. L. (eds.). (1997). *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS).* Oxford University Press.

[19] Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain physiology: II. *Journal of Personality and Social Psychology*, 58(2), 342.

[20] Shackelford, T. K., & Larsen, R. J. (1997). Facial asymmetry as an indicator of psychological, emotional, and physiological distress. *Journal of personality and social psychology*, *72*(2), 456.

[21] W. Han (2014). Multiple Direction Filter as the Feature Map for Image Analysis of Affective Computing. MSc thesis. Imperial College London