Optical Flow Image Analysis of Facial Expressions of Human Emotion – Forensic Applications

Carmen J. Duthoit
Dept. Medical and Molecular Biosciences (MMB)
University of Technology, Sydney
P.O. Box 123 Broadway NSW 2007
+61438317670
Carmen.J.Duthoit@uts.edu.au

T. Sztynda, S.K.L. Lal, B.T. Jap
MMB, Centres of IMS, Engineering and Forensic Science
University of Technology, Sydney
P.O. Box 123 Broadway NSW 2007
+6129514 4157, 1592, 8291
Sara.Lal@uts.edu.au

J.I. Agbinya
Centre for Real-Time Information Networks (CRIN), Engineering
University of Technology, Sydney
P.O. Box 123 Broadway NSW 2007
+61295142495
agbinya@eng.uts.edu.au

ABSTRACT

The objective of this study was to induce emotions in individuals to determine if specific facial movements could be detected and analysed by the optical flow technique. This analysis is in the form of motion vector plots. The procedure ascertains if specific emotions can be defined as a set of facial movements which are common to most people when they experience a particular emotion. 'Emotion vector maps' would then be established for specific emotions. These vector sets could then be applied to automated facial image analysis for security/forensic purposes.

Individuals were videotaped while watching emotion-inducing short films. After the viewing of each short film, volunteers completed a brief self-reporting questionnaire to establish the emotions they experienced. The image sequences were then analysed according to emotion, by using optical flow analysis. Results were statistically analysed.

Trends and analyses are presented in relation to security and video surveillance. Issues and the development of pattern recognition systems applied to human facial images for security purposes are briefly discussed.

Categories and Subject Descriptors

G.4 [**Programming Languages**]: Matlab 7.0.1 – Image processing toolbox. I.4.0 [**Image Processing**] Adobe Photoshop CS2, H.4.1 [**Data Collation, Minor Statistical Analyses**] Excel (Microsoft Office 2003)

General Terms

Experimentation, Human Factors, Measurement, Security.

Keywords

Emotion, facial expression, forensics, surveillance, image analysis, optical flow, physiological measurement.

E-FORENSICS 2008, January 21-23, Adelaide, Australia

DOI 10.4108/e-forensics.2008.2743

1. INTRODUCTION

Emotion is a facet of the human race that has piqued curiosity amongst scientists and psychologists alike, perhaps because of its intangibility, its inability to be quantified. It can most simply be defined as a mental sensation or affection which is distinct from volitions and cognitions [21].

Because of this seemingly vague grasp of the concept of emotions, there have been many attempts to measure or quantify them. In order to do this, manifestations of emotion are the usual target of analysis. This can involve examining facial expressions [10], body movement [19], posture [7], arm movement [20], electromyography [3], physiological changes [18] or even voice characteristics [15].

The most accessible and perhaps simplest method of studying emotion is recording facial expression. This is because it is one manifestation of emotion that is usually easily visible. It is one of the most frequently used methods in everyday life to make an inference about a person's emotional state. This is also the method that was chosen for this research, as it is believed to be one of the more easily observed emotion indicators, and there has not been much recent study or research on the links between emotion and facial expression. It is also almost unanimously agreed amongst researchers that facial expressions are produced or at least effected by the emotional state [9].

2. RESEARCH CONTEXT

In situations where security is of paramount importance, it is very useful to have methods of determining an individual's emotions and whether they are trying to hide their real emotional state, which could be masking criminal behaviour. If one can understand and interpret an individual's emotion, such knowledge can be applied to the interpretation of behaviour of suspicious individuals in many settings. Every day, new methods are being developed to improve security both locally and internationally to increase the chance of catching the 'bad guys'. For example, over several years, a method called Screening Passengers by Observation Techniques (SPOT) has been used in airports in both Israel and the United States [16]. This involves using techniques that were originally used to identify possible drug couriers. Police officers observe traveller's signs of deception, stress or disorientation by looking at facial expressions, body language and listening to vocal cues. Individuals considered suspicious are identified and then referred for intense screening. This method was prompted by the 2001 September 11 attacks in the United States, in which thousands of people lost their lives and many more were injured. If such a method had been used prior to this event, it may have been possible to single out the plane hijackers by this technique and prevent this catastrophic incident.

3. METHODS

3.1 Choice of Methodology

3.1.1 Emotion Induction

When undertaking emotion research, there are many issues to address. The first, and perhaps most troublesome, is how to induce emotion. There have been many methods proposed and used in the past by different researchers. Gerrards-Hesse, *et al.* [12] proposed five main groups of 'mood induction procedures' {note: the referenced study used emotion and mood as synonyms, which these authors do not. Therefore for the sake of accuracy, the phrase 'mood induction procedures' will be replaced with that of emotion inducing procedure (EIP)}. The five groups of EIPs [12] are:

Group 1 with a foundation in the free mental generation of emotional condition (such as hypnosis and imagination).

Group 2 formulated on the presentation of emotion-inducing matter (such as film, story and music).

Group 3 founded on guided mental generation of emotional states (such as those mentioned in the previous point but with the addition of explicit instruction to get into the suggested emotional state).

Group 4 based on the production of need-related emotional states (such as production of social interaction and false feedback).

Group 5 based on the creation of emotionally relevant physiological states (such as drugs and facial expression [which they believe will induce emotion]).

The study then went on to discuss the effectiveness of each type of EIP, reviewing over 200 emotion induction studies and came to the conclusion that the film/story EIP was the most effective for inducing both elation and depression (or 'positive' and 'negative' emotions). Other benefits to this method, according to Gross and Levenson [14], include that it is very easily standardised, and is a dynamic rather than static process. This is important because it is also representative of how emotions are evoked in real life, i.e. they are usually produced by dynamic external stimuli.

It was decided to use short films to induce emotion. This still allows an individual to get emotionally involved in a situation, whilst eliminating the need for any editing procedures for feature-length films (as in Gross and Levenson's study [12]), because it would not be practical to show numerous feature-length films, and would be very difficult to find one film that would induce the variety of emotions that we aimed to examine in this study. Such a technique would also involve recording significant amounts of data that would have no relevance to the research (a film does not constantly evoke emotions) and would result in a large amount of data to analyse. Out of all the emotion induction methods, the screening of emotionally laden films is considered to be the least problematic [22].

3.1.2 Image Analysis

There are many elements considered when choosing a method for facial expression analysis. An important one is that the method must be applicable to non-rigid motion - like that of facial expressions.

The core of the analysis is based on facial expressions produced by the instigated emotion/s as mentioned in the previous section. There are many ways to analyse facial expressions, for example the Facial Action Coding System (FACS) [5], Facial Expression Movement System (FACEM) [4], optical flow [1], holistic special analysis (otherwise known as image subtraction) [11], among others.

3.1.2.1 Optical Flow

The method chosen for the analysis of facial expressions was the optical flow technique [see 1 & 11]. This method involved assessing the magnitude and direction of facial motion. In recent years, optical flow has emerged as a useful tool in the analysis and tracking of motion features in video sequences. The traditional approach assumes the motion between two image frames at the pixel level. The motion of an object in three-dimensions (3D) is projected into a two-dimensional (2D) plane. It is assumed the type of motion sought is distributed over a sequence of several neighbouring frames. In practice this assumption is reasonable as the motion is gradually spread out across several frames. Although the large motion of vehicles and human beings is easier to detect, the morphological deformation of a human body organ is harder to capture. For this work, it was assumed there is a brightness constancy across neighbouring video frames. Hence the brightness of neighbouring image pixels does not change over time. Therefore for two image frames the following applies:

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t)$$

By using Taylor series expansion, small motion of pixels in two neighbouring images is given by the expression:

$$I(x, y, t) = I(x, y, t) + \Delta x \frac{\partial I}{\partial x} + \Delta y \frac{\partial I}{\partial y} + \Delta t \frac{\partial I}{\partial t}$$

The brightness constancy assumption results to:

$$0 = \Delta x \frac{\partial I}{\partial x} + \Delta y \frac{\partial I}{\partial y} + \Delta t \frac{\partial I}{\partial t} = \Delta x I_x + \Delta y I_y + \Delta t I_t$$

Or $uI_x + vI_y + I_t = 0; \quad u = \frac{\Delta x}{\Delta t}; \quad v = \frac{\Delta y}{\Delta t}$

Hence the optical flow expression in the u and v space is an equation of a line:

$$v = u \frac{I_x}{I_y} + \frac{I_t}{I_y}$$

For the work reported in this paper the Lucas and Kanade formulation [17] of this expression has been used, which defines motion energy functions, the Laplacian of the image, and assumes that two neighbouring image pixels move at the same velocity. A Matlab program was written to implement the Lucas and Kanade optical flow algorithm. Where a reference video frame and a video frame of the same size are analysed to estimate the quantity of motion between the two frames. This motion is presented as a vector space situated within the space of the image frame. In addition, to estimate the directions of the dominant motion vectors, the results are presented as a compass that depicts the vectors in a 360 degrees circular plane.

In Figure 1 an example of an optical flow analysis can be seen, where the 'neutral face' seen in image A is compared to an image of the 'expressive face' (image B) creating a vector map (quiver diagram) - Image C. The arrows are 'flow vectors', with their length being proportional to the velocity of the expression [1].

Image A

Image B

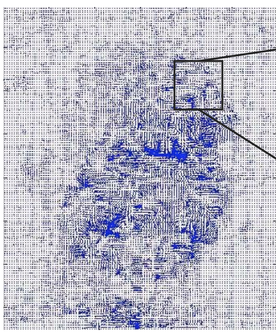


Image C

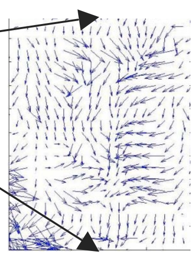


Image D

Figure 1 Image Analysis by Optical Flow (quiver diagram)

The base 'neutral face' image A[1] on the left is compared to the 'expressive face' image B and the vector map image C is created. Image D is an enlarged portion of the complete vector map, or quiver diagram, showing individual vectors, which are too difficult to see in image C.

3.2 Experimental Process

3.2.1 Volunteers

Subjects were recruited from the general public by email, posters online recruitment and word of mouth, with the only limitations being age (18-65 years), and English as language most often spoken at home. Age was considered important as emotional expression may not be totally developed before the age of 18 years (it is not known at what point an individuals emotional development is complete), and after 65, there may be issues regarding deterioration of an individual's memory, and also illnesses or conditions such as stroke, which can result in paralysis of part of the face and prevent the completion of a facial expression. The language most often spoken at home goes towards an individual's understanding of the short films.

[1]Permission obtained to reproduce photo of the individual in Figure 1

3.2.2 Laboratory Set-up

The experiment was set-up in a room 3 by 5 m. The DVD/TV equipment was arranged at the short end of the room, with the TV being placed on two joined tables, and the DVD player to one side. A camera was located on a tripod behind and slightly above the TV, with the lens facing a chair positioned approximately 1.5m from the TV. A second camera was placed 70 cm to the left side of the chair, at head height, aimed to capture the participant's left profile.

3.2.3 Physiological Monitoring

Blood pressure and heart rate measurements were performed. These readings were taken with the OMRON Digital Blood Pressure Monitor, after the administration of the initial questionnaires in order to allow for a period of rest from when the participant arrived (to reduce possible interference of physical activity on resting heart rate or blood pressure).

3.2.4 Questionnaires

After emotion has been induced, before even methods of recording are considered, it was vital to assess the success of the emotion induction procedure, or to perform 'manipulation checks' [19]. This was done by self-assessment questionnaires (a self-report of emotional state). The 'Duthoit' questionnaire was specifically designed for this research, being a. brief film-specific self-reporting assessment based on that described and used in Gross and Levenson [14]. The questions asked the participant how they felt at predetermined moments in the film (e.g. when the police shot the robber). This was done so that the recorded facial expressions could be synchronized with the emotion being experienced at that time. This process was completed for all films.

The participants were also asked to fill out two questionnaires before the film screening (the Lifestyle Feedback Questionnaire [8] and the Emotion Regulation Questionnaire, or ERQ [11]). These were administered to determine the subject's suitability for the study.

3.2.5 Experimental Process

A total of 30 volunteers offered their services, most watched 5 short films each. All volunteers had their blood pressure taken, before the commencement of the study when they answered a lifestyle questionnaire.

The participants were seated on a chair approximately 1.5m away from the TV screen, directly facing it, and instructed to look at the TV. This process assisted in determining the link between emotion and associated physiological changes.

The participants were given the remote control and instructed to adjust the volume at any time to a comfortable level during the films. The cameras were then turned on, the film started and the researcher left the room.

The short films were shown in a random sequence with the purpose of negating order effect. The initial seven films chosen were: *Lamb, Tickler, The Thing in the Roof, Late Night Shopper, Rattus Pistofficus, Desy* and *Too Sunny, Too Cold* from various years of the Australian Tropfest Short Film Festival DVDs (permission granted). After the first ten participants, two films (*Rattus Pistofficus* and *Too Sunny Too Cold*) were excluded because they were not successful at inducing emotions, this also shorten the experimental process making it less onerous for volunteers.

After the completion of each film, the cameras were turned off. Blood pressure and heart rate measurements were retaken. The participant was then asked to fill out one of the 'Duthoit' questionnaires.

4. RESULTS

4.1 Data Collection and Collation

Every volunteer had 5 recordings of their face taken as they watched each of the films. Each such digital video recording was then examined in light of the film-specific questionnaire, and at least 2 still images extracted as they experienced their 'self-reported' emotion, such that it included the start, climax and end of the specific emotion represented.

Then the neutral and the emotional facial images were resized in Adobe Photoshop to a specific number of pixels (required for exact image comparisons), so that they could be recognised by Matlab Image Processing Toolbox.

4.2 Image Analysis

Optical flow analysis on each of the images (comparised with the baseline image of that particular person taken before the film was started) resulted in vector maps, an example of which is shown in Figure 1. The vectors (arrows) as shown in the figure represent angles and amplitude of the summed facial movements associated with a particular emotion – 'Amusement' as depicted in this map. The quiver vector map was converted using Matlab to a compass function vector diagram showing the summation of the directional vectors (Figure 2), creating representation of the whole face from the quiver diagram. This compass diagram summed all the vectors from the quiver diagram and created a determination of the total directional vectors from the centre of origin and seen in segments of a 'navigational compass'. It shows the major directions and velocities of facial movement.

This compass is not to be confused with a human face divided into segments, nor is this centre origin equivalent to the centre of the face. It simply represents the sum of all directional vectors from the centre origin.

The compass diagram (360 degrees) is divided into 12 segments each of 30 degrees, and represents the total directional facial movement from the original neutral image. Vectors radiate from the origin (the centre of the compass – zero position). These data were also tabulated and the vectors summed for each 30 degree segment and the top 5 (magnitude) directional segments determined to ascertain the most prominent facial vector flow.

The numerical values gave little information, and determination of mean and standard deviation was not possible as the specific segments, although frequently recurring, did not commonly group together. Therefore it was decided to colour code the specific 30 degree segments and try to visually determine any patterning, (demonstrated in Figures 3, compass diagrams for pictures (PIC1, 2 and, 3).

Comparing the compass diagrams from figure 3, it can be seen that in this particular volunteer, the expression of 'amusement' consistently produces movement in the orange (60-90 degrees), red (90-120 degrees) and light blue (180-210 degrees) sectors. There are a number of segments that are also consistently inactive.

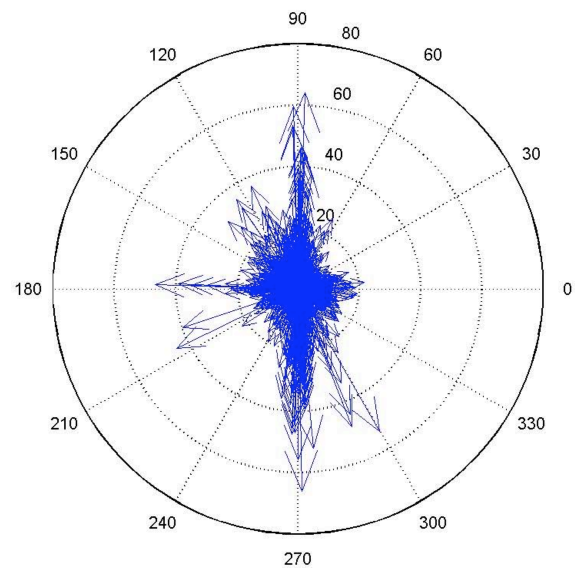


Figure 2 Optical Flow - Compass Function of Vector Diagram of Amusement

The quiver vector map was converted using Matlab to a compass diagram showing the summation of the directional vectors. This compass diagram shows the major directions and velocities of facial movement.

When summarising this information for a number of individuals experiencing amusement in different situations (shown in figure 4) it can be seen that although between individuals there does not appear to be much commonality in the sectors that are active during the emotion, there is consistency in the sectors that are inactive. This is demonstrated by the fact that for all 3 individuals shown, there is lack movement in the grey (0-90 degrees) and yellow (120-150 degrees) sectors.

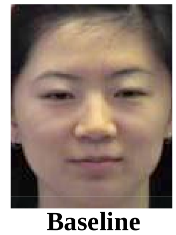
5. DISCUSSION

From these preliminary results that it is possible to identify consistencies in an optical flow compass diagram for an individual experiencing the same emotion in different circumstances. The example presented all three participants for whom 'amusement' was analysed demonstrated similarities each time they experienced the emotion. These similarities were seen not in the active segments of the compass diagrams as expected, but in the inactive segments.

This pilot study bodes well and the data require further in-depth investigation, such as principal component analysis (PCA). There were a number of complicating issues which arose; such as the way in which the volunteers answered the questionnaires (video sequence could only be selected if one emotion was experienced). This resulted in rejection of a large amount of the data. However, trends can be seen which indicate that once the protocol is further refined greater reproducibility will be possible on which further developmental work can be based.

Figure 3

	1	2	3	4	5
pic 1	102.27	192.33	346.92	76.72	165.04
pic 2	346.64	101.80	192.32	46.14	76.70
pic 3	165.85	193.69	254.92	76.75	102.20



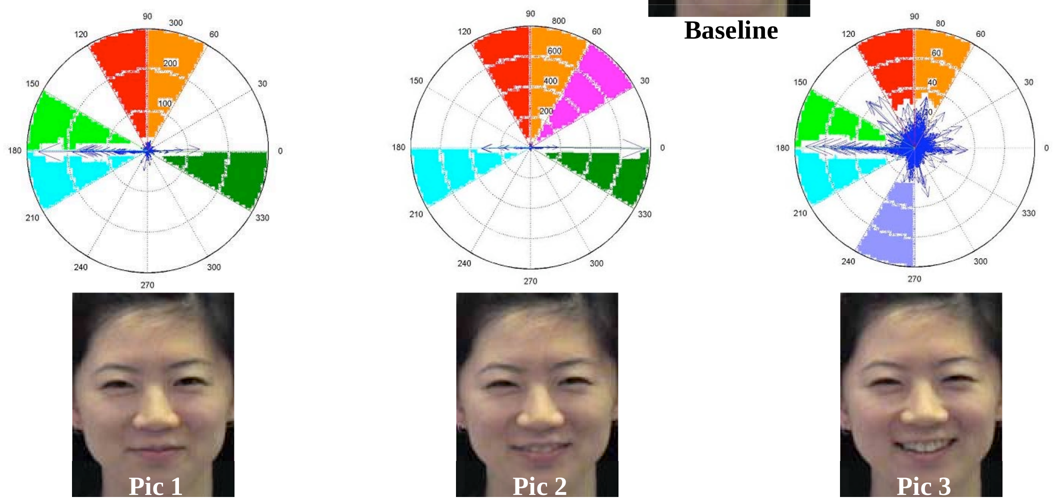


Figure 3: Example of subject experiencing 'amusement'

Colour coded compass diagrams of major movement vectors for 'amusement' for a volunteer. The table enumerates the exact angles which predominate in the 5 most prominent sectors. Pictures[2] (Pic1 to 3) show the development of amusement attributed to the compass diagram above. [2] Permission obtained to reproduce photo of the individual in Figure 1

Figure 4

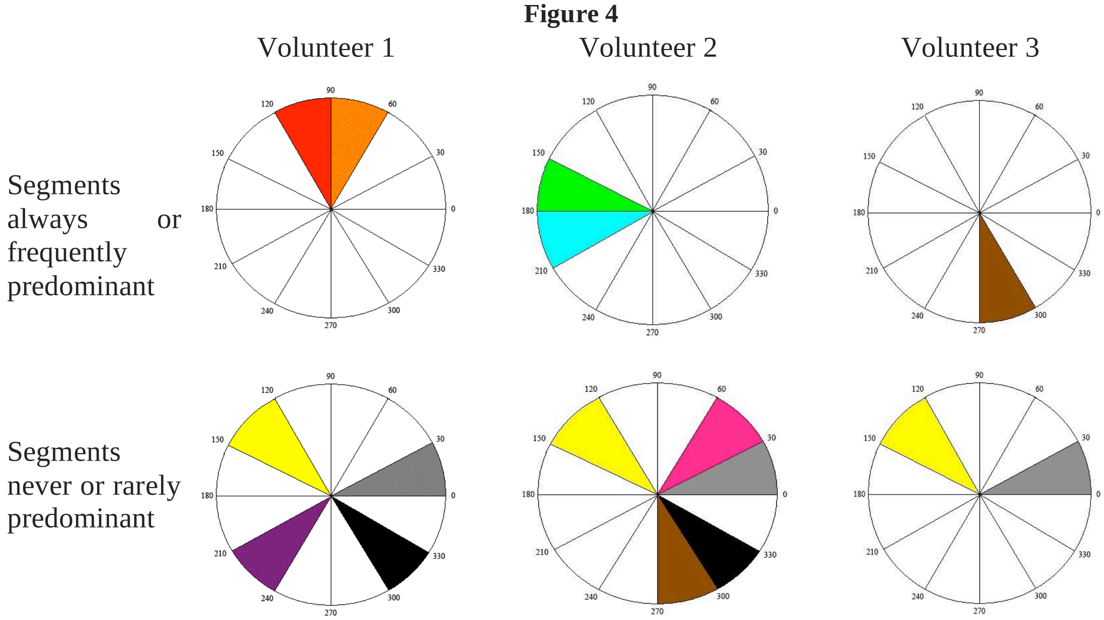


Figure 4: Summary of common compass segments in amusement for 3 different individuals

The top line of summary diagrams represents the sectors of the compass diagram commonly active for each individual for the emotion 'amusement' and the bottom shows the sectors commonly inactive for amusement in these same individuals. We can see that whereas there is no consistency between individuals in the active segments, all 3 individuals show absence of movement in the yellow and grey sectors.

The eventual aim is to be able to produce a system that allows for the recognition of facial expressions/emotions without the need for specialised training such as with FACS, being personnel based is expensive and time consuming. This system may then be utilised in security sensitive situations, such as in airports, where is may be beneficial to identify an individual that appears 'suspicious'. It is also hoped that automation of this technique may be achieved. It is also possible to develop this research and integrate it into a system along with face recognition processes [2] and face detection software [5] allowing for use on surveillance footage and application in forensic analyses.

6. ACKNOWLEDGMENTS

Institutional ethics approval obtained.

Research support from the faculties of Science and Engineering, UTS.

Frank Zhang for technical support.

7. REFERENCES

[1] Bartlett, M.S., Hager, J.C., Ekman, P. & Sejnowski, T.J. 1999, 'Measuring facial expressions by computer image analysis', Psychophysiology, vol. 36, pp. 253-263.

[2] Beymer, D. and Poggio, T. (1995), 'Face Recognition From One Example View', *Fifth International Conference on Computer Vision*, ed. -, The Institute of Electrical and Electronics Engineers, Inc., Cambridge, Massachusetts, pp. 500-507

[3] Burriss, L., Powell, D.A. & White, J. 2007, 'Psychophysiological and subjective indices of emotion as a function of age and gender', Cognition & Emotion, vol. 21, no. 1, pp. 182 - 210.

[4] Byrne, T., Henskens, F., Johnston, P. & Katsikitis, M. 2003, 'FaceXpress: An Integrated Software Suite for Facial Emotion Stimulus Manipulation and Facial Measurement', Methods of Psychological Research Online, vol. 8, no. 3, pp. 97-111.

[5] Chen, Q., Wu, H. and Yachida, M. (1995), 'Face Detection by Fuzzy Pattern Matching', *Fifth International Conference on Computer Vision*, ed. -, The Institute of Electrical and Electronics Engineers, Inc., Cambridge, Massachusetts, pp. 591-596

[6] Chesney, M.A., Ekman, P., Friesen, W.V., Black, G.W. & Hecker, M.H.L. 1990, 'Type A Behavior Patter: Facial Behavior and Speech Components', Psychosomatic Medicine, vol. 53, pp. 307-319.

[7] Coulson, M. 2004, 'Attributing emotion to static body postures: Recognition accuracy, confusions and viewpoint dependance', Journal of Nonverbal behaviour, vol. 28, no. 2, pp. 117-139.

[8] Craig, A., Hancock, K. & Craig, M. 1996, 'The lifestyle appraisal questionnaire: A comprehensive assessment of health and stress', Psychology and Health, vol. 11, pp. 331-343.

[9] Dimberg, U., Thunberg, M. & Grunedal, S. 2002, 'Facial reactions to emotional stimuli: Automatically controlled emotional responses', Cognition & Emotion, vol. 16, no. 4, pp. 449 - 471.

[10] Ekman, P. 1992, 'FACIAL EXPRESSIONS OF EMOTION: New Findings, New Questions', Psychological Science, vol. 3, no. 1, pp. 34-38.

[11] Essa, I.A. & Pentland, A.P. 1995, 'Coding, Analysis, Interpretation, and Recognition of Facial Expressions', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. April 1995, pp. 1-23.

[12] Gerrards-Hesse, A., Spies, K. & Hesse, F.W. 1994, 'Experimental inductions of emotional states and their effectiveness: A review', British Journal of Psychology, vol. 85, pp. 55-78.

[13] Gross, J.J. & John, O.P. 2003, 'Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being', Journal of Personality and Social Psychology, vol. 85, pp. 348-362.

[14] Gross, J.J. & Levenson, R.W. 1995, 'Emotion Elicitation Using Films', Cognition & Emotion, vol. 9, no. 1, pp. 87-108.

[15] Laukka, P., Juslin, P. & Bresin, R. 2005, 'A dimensional approach to vocal expression of emotion', Cognition & Emotion, vol. 19, no. 5, pp. 633 - 653.

[16] Lipton, E. 2006, 'Faces of Terror?' New York Times Upfront, vol. 139, no. 2, pp. 8-9.

[17] Lucas, B.D. and Kanade, T. (1981), 'An iterative image registration technique with an application to stereo vision', Proceedings IJCAI, Vancouver, Canada, pp. 674-679.

[18] Mera, K. & Ichimura, T. 2004, 'Emotion Analyzing Method Using Physiological State', in, Knowledge-Based Intelligent Information and Engineering Systems, Springer Berlin / Heidelberg, pp. 195-201.

[19] Montepare, J., Koff, E., Zaitchik, D. & Albert, M. 1999, 'The use of body movements and gestures as cues to emotions in younger and older adults', Journal of Nonverbal behaviour, vol. 23, no. 2, pp. 133-152.

[20] Polivy, J. 1981, 'On the Induction of Emotion in the Laboratory: Discrete Moods or Multiple Affect States?' Journal of Personality and Social Psychology, vol. 41, no. 4, pp. 803-817.

[21] Toates, F. 2001, Biological Psychology An Integrative Approach, Pearson Education Limited, Essex.

[22] Yeasin, M., Bullot, B. & Sharma, R. 2006, 'Recognition of Facial Expressions and Measurement of Levels of Interest From Video', IEEE Transactions on Multimedia, vol. 8, no. 3, pp. 500-508.