Spontaneous Facial Expression Analysis Using Optical Flow

Lina Sidavong, Sara Lal, Tamara Sztynda
School of Life Sciences
University of Technology Sydney, Australia
Lina.Sidavong@uts.edu.au Sara.Lal@uts.edu.au Tamara.Sztynda@uts.edu.au

Abstract—**Investigation of emotions manifested through facial expressions has valuable applications in predictive behavioural studies. This has piqued interest towards developing intelligent visual surveillance using facial expression analysis coupled with Closed Circuit Television (CCTV). However, a facial recognition program tailored to evaluating facial behaviour for forensic and security purposes can be met if patterns of emotions in general can be detected. The present study assesses whether emotional expression derived from frontal or profile views of the face can be used to determine differences between three emotions: Amusement, Sadness and Fear using the optical flow technique. Analysis was in the form of emotion maps constructed from feature vectors obtained from using the Lucas-Kanade implementation of optical flow. These feature vectors were selected as inputs for classification. It was anticipated that the findings would assist in improving the optical flow algorithm for feature extraction. However, further data analyses are necessary to confirm if different types of emotion can be identified clearly using optical flow or other such techniques.**

Keywords—emotion induction; Lucas Kanade Optical Flow; spontaneous emotional expression; facial recognition.

I. INTRODUCTION

Heightened concerns in security and public safety have led to the increased use of surveillance cameras such as Closed Circuit Television (CCTV) systems to monitor the behavior of individuals in efforts to reduce and deter crime [1]. Despite this growth, much controversy surrounds the efficacy of such operations. There is a significant cognitive demand placed on the human operators who monitor CCTV systems [2]. The vigilance required of human observers dramatically declines over time, resulting in overlooking crucial information that would alert them to an emerging incident [3].

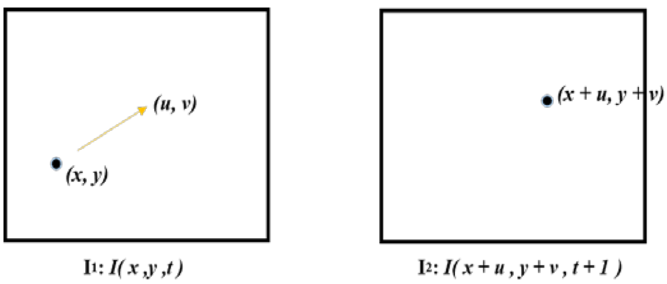
Aligning with the objectives to enhance proactive security, an approach may be to impart intelligence to surveillance systems, such as developing automatic systems for recognising facial expressions. Application of facial expression recognition in law enforcement and counter terrorism has profound consequences for improving security. It can play a role in crime prevention as expressions are valid reflections of a person's mental state and allow the prediction of future behaviours [4]. The idea is to develop an algorithm for automated recognition of emotions by analysing facial expressions through CCTV and other video surveillance system images. This automation could serve as an early warning system because it can detect a person's expression of intentions for causing harm.

A. *Automatic Facial Expression Recognition*

An automatic facial expression recognition (AFER) system is defined by its ability to effectively measure and classify expressions of emotion. The designed configuration fundamental to most systems comprises: face detection, feature extraction and expression classification [5]. Face detection initiates the recognition processes by locating the face within a scene and often requires pre-processing to achieve a pure isolate of the facial image. Feature extraction transforms high dimensional data into low dimensional features as inputs into classifiers. Not only does removing reductant features make the system computationally efficient, the classification of new instances is more accurate. In saying so, the feature extraction and its selection process is critical for predicting the success of an AFER system, as inadequate feature selection may results in failure of accurate recognition.

B. *Optical Flow*

The optical flow technique has been applied to expression recognition to extract dynamic facial movement image sequences [6-8]. The approach extracts motion from the changes in movement patterns of the skin to discriminate facial displays from successive images. The differential method of optical flow is based on a local Taylor series approximation using the brightness constancy constraint as the common starting point for optical flow estimation. It assumes that pixel intensities are translated from one frame to the next. The explanations of optical flow are as follow:



Where $I(x, y, t)$ is the intensity of a pixel at location (x, y) in image I_1 at time t and is offset by the flow (u, v) at time $t+1$ in image I_2. The brightness constancy is given by:

$$I(x, y, t) = I(x + u, y + v, t + 1) \quad (1)$$

$$I(x, y, t) = I(x, y, t) + u\frac{\delta I}{\delta x} + v\frac{\delta I}{\delta y} + \frac{\delta I}{\delta t} \quad (2)$$

$$0 = u\frac{\delta I}{\delta x} + v\frac{\delta I}{\delta y} + I\frac{\delta I}{\delta t} \quad (3)$$

The partial derivatives can be substituted for the intensity derivatives:

$$Ix = \frac{\delta I}{\delta x}, \qquad Iv = \frac{\delta I}{\delta y}, \qquad \text{and } It = \frac{\delta I}{\delta t}$$

$$I_x u + I_y v + It = 0 \quad (4)$$

The equation can be rewritten as:

$$\nabla I.\vec{v} = -It \quad (5)$$

Where ∇I is the spatial intensity gradient and $\vec{v}$ is the image velocity or optical flow of pixel (x, y) at time t. Although the brightness constancy equation defines the gradient of the moving object, the boundaries of motion remain obscure. Thus additional constraints for flow determination must be introduced [9].

C. Lucas-Kanade Method

The Lucas-Kanade framework assumes that motion is smooth within a local image neighbourhood [10]. The method obtains the optical flow velocities using the brightness constancy equation for a 3x3 neighbourhood around the pixel using the least squares criterion [11]. As such, the optimal value of flow velocities can be achieved by minimizing this error function with respect to u and v.

Whilst humans recognise facial expressions effortlessly without delay, automatic recognition of facial expressions by machine remains a challenge. Most of the existing studies limit facial expression recognition to frontal facial image analysis. Systems amended to real world applications should allow for faces to be reliably decoded at various angles not limited to full features of the whole face, which may not be continuously available on camera [12]. Whilst expression classification works reasonably well for posed expressions, their performance on spontaneous (impromptu) expressions drops quite dramatically. Given that subtle expressions occur in real-life more frequently, the approaches may fall short in generalising real world behaviour.

The purpose of the study was to determine if specific patterns and trends in facial movements could be detected from the frontal and profile views of the face during an emotional expression. It was best seen to assign emotions the simplest to manifest (i.e. amusement, sadness and fear) so a prototype algorithm for recognising these emotions can be written.

Electrodermal activity (EDA) patterns were measured along with self-reports to authenticate the presence of emotion. The Lucas-Kanade optical flow technique was implemented to define pattern changes in facial movements most common to people when they experience an emotion.

II. METHODOLOGY

A. Subjects

The study was conducted on 65 (n) healthy volunteers (M= 19, F = 46) aged between 18-34 years (mean = 24.2 years, SD = 4.31). This study had human ethics approval from the University of Technology Sydney, Human Research Ethic Committee (UTS-HREC) and the subjects gave informed consent to participate in the study.

B. Emotion Induction

Individuals were videotaped while watching three emotion inducing short films. The films were shown with the intention of inducing one of three emotions: amusement, sadness or fear. The films shown to elicit amusement, sadness and fear were a series of Never Say No to Panda clips (2010) created by Advantage Marketing and Advertising company, Last Minutes with Oden (Directed by Eliot Rausch, 2009) and Lights Out (Directed by David F. Sandberg, 2013) respectively. Permission for use of the short films was granted by the film directors and the Never Say No to Panda clips (2010) was acquired from agency website.

Skin conductance (SC) was measured by attaching two Ag/AgCl electrodes to the medial phalange of the middle and index fingers of the non-dominant hand. Observed arousal peaks correlating to self-reports of emotion (using a nine-point Likert scale questionnaire) were used to establish the type and extent of emotion felt. This allowed for extraction of still facial images representative of neutral “baseline” and “peak of emotion” expressions, for each subject from the front, profile left and profile right sides of the face (Fig. 1).

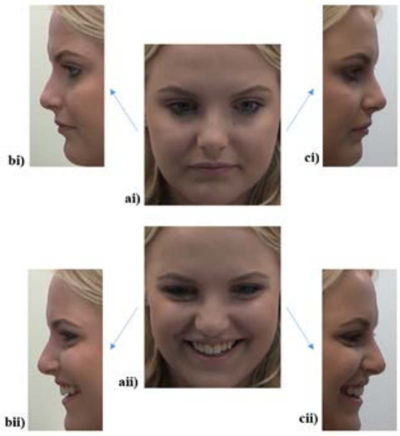


Figure 1. *Image sets consisting of facial images representative of (i) neutral and (ii) peak of emotional expression for the (a) frontal, (b) profile left and (c) profile right sides. (Permission granted for use of images).*

C. *Image Analysis*

Optical flow analysis was performed on the image sets using MATLAB R2013b to quantify the magnitude and direction of facial activity between neutral and peak emotional states. A quiver plot was constructed depicting the flow of facial movement as directional vectors. A compass graph was produced which summed all the vectors onto a 360° plane, divided into 12 sectors of 30° (see Fig.2).

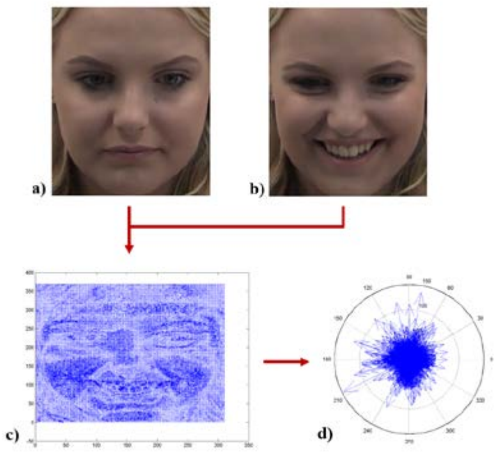


Figure 2. *Optical flow computed motion activity between still facial images representative of (a) neutral and (b) peak of emotional expression onto a quiver plot (c). The computed motion in the form of velocity vectors was rearranged onto (d) a compass graph as arrows emanating from the central axis. (Permission granted for use of images).*

Optical flow output creates twelve feature vectors by calculating the average vector magnitude for each sector. These feature vectors were used to establish emotion vector maps for specific emotions. The emotion map is an adapted form of the compass graph whereby each 30° sector of the map was colour coded for visual ease of quick segment comparisons between emotions (Fig. 3).

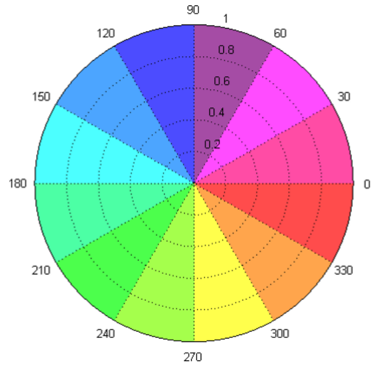


Figure 3. Colour coding for emotion maps. *An adapted form of the compass graph starting with 0° on the right and moving counter clockwise for increasing angles is divided into twelve even sectors. Each sector colour coded for visual ease of analysis.*

III. RESULTS

An exploratory analysis of optical flow data was conducted with an aim to detect patterns and trends to differentiate between emotional facial expressions: amusement, sadness and fear. The distribution of emotion class (by number of volunteers responding) in the acquired dataset are displayed in Table 1.

TABLE I. DISTRIBUTION OF EMOTIONS IN THE DATASET

	Emotion Class (number of responding volunteers where n = 65)		
	Amusement	***Sadness***	***Fear***
Frontal	63	42	34
Profile Right	60	38	34
Profile Left	61	38	33

Analysis involved mapping feature vectors of facial activity for an expression onto emotion maps according to the level of activity. Data were ranked to extract the highest and lowest areas of movement. The sectors generating the four relative frequencies for the highest and lowest division were transferred onto emotion maps. Fig. 4 shows the distribution of the sectors which exhibited the lowest facial activity for the emotional expression of amusement, sadness and fear from the frontal and profile views. Where the radial distance (r) of the sectors from the centre origin correlate to the magnitude value of feature vectors.

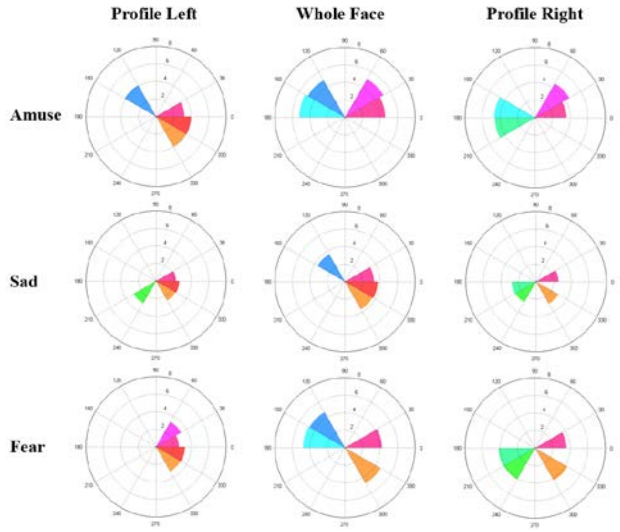


Figure 4. Emotion maps representing lower division of movements. *The coloured segments represent the average feature vector values in the lower division of movement commonly expressed by subjects during the emotions: amusement, sadness and fear.*

When data were sorted to extract the highest and lowest areas of movement, we observed consistent trends and patterns of activity in the higher division group. However, variation in movements was seen in regions of lower divisions of activity.

Statistical analysis was applied to determine whether commons patterns of movement differed in intensity. A Repeated Measures analysis of variance (RM-ANOVA) with a Greenhouse-Geisser correction was carried out to compare the means of common sectors in the lower areas of movement to determine whether there was a statistically significant difference in the magnitude of facial activity. *Post hoc* using the Bonferroni correction revealed a statistically significant difference ($p < 0.05$) in activity between the expression of amusement and sadness in sectors 1 (0°-30°) and 5 (120°-150°) for frontal images, amusement and sadness in sectors 7 (180°-210°) and 8 (210°-240°) for the lowest area of movement (TABLE II). Activity was greater on the profile right than profile left sides. No significant differences ($p > 0.05$) were found for facial activity between the three emotions from the left profile view.

TABLE II. RESULT OF POST HOC COMPARISON FOR MEAN AREA OF MOVEMENT

	Sectors	**(I) Emotion**	**(J) Emotion**	**Mean difference (I-J) (*)**	**Std.error**	***P* value**	**95% confidence interval for difference**	
							Lower bound	**Upper bound**
Frontal	**1**	**A**	**S**	1.207	0.371	0.010	0.252	2.161
	5	**A**	**S**	1.248	0.404	0.014	0.216	2.279
Profile Right	**7**	**A**	**S**	1.259	0.419	0.02	0.174	2.345
	8	**A**	**S**	1.343	0.394	0.008	0.322	2.365

Note: The mean difference is significant at the 0.05 level; A = Amusement; S = Sadness

To test whether the trends and pattern of activity observed could be used to classify emotions. Feature vectors depicting the lowest area of movement were selected as inputs as well as using the all feature vectors offered by optical flow. Three classifiers were employed from the Classification Learner application in MATLAB: linear Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), and K-Nearest Neighbor (KNN). The total accuracy of classifying all three emotions: amusement, sadness and fear using all feature vectors from the optical flow and the subset of feature vectors representing the lower division of movement are shown in TABLE III and TABLE IV respectively.

Overall, the frontal facial images yielded slightly superior accuracy than the facial images obtained from the left and right profile views. SVM and KNN classifiers showed consistency in recognition between the two sets of feature vectors. The LDA classifier boosted the accuracy of the profile views using the subset of feature vectors. Accuracy of classifying individual emotions using the linear SVM is shown in TABLE V. It is observed that amusement has the highest recognition followed by sadness. Recognition of the expression of fear was remarkably low, particularly on the profile left side which had a negative impact on the overall accuracy.

TABLE III. TOTAL PERCENTAGE (%) ACCURACY OF CLASSIFYING EMOTIONS USING ALL FEATURE OUTPUTS PERCENTAGE

	Classifiers		
	SVM	***LDA***	***KNN***
Frontal	60	59.3	57.9
Profile Right	57.6	45.5	56.1
Profile Left	50.8	47.7	49.2

Note: SVM = Support Vector Machine; LDA = Linear Discriminant Analysis; KNN = K- Nearest Neighbors

TABLE IV. TOTAL PERCNTAGE (%) ACCURACY OF CLASSIFYING EMOTIONS USING LOWEST FEATURE MAGNITUDE OUTPUT

	Classifiers		
	SVM	***LDA***	***KNN***
Frontal	60	60.7	52.9
Profile Right	55.3	59.1	53.8
Profile Left	53.8	51.5	50.8

Note: SVM = Support Vector Machine; LDA = Linear Discriminant Analysis; KNN = K- Nearest Neighbors

TABLE V. PERCENTAGE (%) ACCURAGE OF CLASSIFYING INDIVIDUAL EMOTIONS USING LINEAR SVM

	Emotion Class	**All 12 Features**	**Lowest Feature Magnitude**
Frontal	**A**	73	73
	S	49	52
	F	45	44
Profile Right	**A**	65	64
	S	48	51
	F	50	55
Profile Left	**A**	56	61
	S	48	47
	F	20	32

Note: A = Amusement; S = Sadness; F = Fear

IV. DISCUSSION

We present preliminary results from the application of an optical flow analysis of spontaneous emotional facial expressions. The study used images extracted from video recordings of subjects watching three emotion inducing short films. Optical flow analysis produced feature vectors representing mean global velocity vectors of facial movement constrained within 30° segments. Features were selected that depicted higher and lower areas of movement to determine whether there were differences in facial activity between them. We were able to construct emotion maps for the expressions of amusement, sadness and fear across the whole face and profile views for observable differences and similarities.

When summarising this information derived from the emotion vector maps of amusement, sadness and fear, we observed greater magnitude of activity from frontal facial images than profile views with amusement having an overall greater intensity of movement. This was expected due to the decrease in visibility of the morphological changes that occur on the surface area of the face. Visual inspection of emotion maps showed greater variations in patterns of movement in regions of lower activity for all three emotions. This may be attributed to the action of large, dominant facial muscles on the lesser muscles or by independent mimetic movements of muscles characteristic to each emotion.

Initial findings showed that emotion maps of areas that appeared most active in subjects experiencing the emotions of amusement, sadness and fear could not be differentiated. Comparisons between them showed recognition was not negatively affected by the subset of feature vectors derived from the emotion maps. In fact, the SVM classifier could achieve up to 60% using both and the LDA classifier 60.7% recognition accuracy. However, classification of individual emotions showed recognition of amusement was much higher in comparison to the recognition of fear which may have affected the overall accuracy rate. Nonetheless, we contend that feature vectors depicting low activity regions show potential for differentiating between affective states being related to the range of each individual's response as a mitigating factor and warrants further investigation.

V. CONCLUSION

This work advances the research on facial expression analysis for potential applications in forensic and security surveillance. From this preliminary assessment of spontaneous emotional facial images, optical flow analysis has demonstrated capabilities for discriminating between emotional facial expressions. However, further data analyses are required to confirm if different types of emotion can be identified clearly using optical flow or other such techniques. Due to the ease of utility in CCTV, there is tremendous potential for this program to be used in counter-terrorism, community surveillance and interrogation settings.

ACKNOWLEDGMENT

This research is supported by an Australian Government Research Training Program Scholarship and UTS Science Faculty Research funds. Thanks to Dr. Budi Jap who proposed the feature extraction methods.

REFERENCES

[1] D. Wilson and A. Sutton, "Open-Street CCTV in Australia," *Australian Institute of Criminology, Trends and Issues in Crime and Criminal Justice*, 2003.

[2] F. M. Donald, & Donald, C.H.M. , "Task disengagement and implications for vigilance performance in CCTV surveillance" *Cognition Technology and Work,* vol. 17, pp. 121-130, 2015.

[3] A. Isnard, "Can Surveillance Cameras be Successful in Preventing Crime and Controlling Anti-Social Behaviours," presented at the The Character, Impact and Prevention of Crime in Regional Australia Conference, Townsville, Queensland Australia, 2001.

[4] S. J. Ahn, J. Bailenson, J. Fox, and M. Jabon, "Using Automated Facial Expression Analysis for Emotion and Behavior Prediction," In: *Handbook of Emotions and Mass Media*, K. Doeveling, C. von Scheve, and E. A. Konjin, (Eds.), London/New York: Routledge, 2010, pp. 349-369.

[5] C. P. Sumanthi, Santhanam, T. & Mahadevi, M. , "Automatic Facial Expression Analysis: A Survey," *International Journal of Computer Science & Engineering Survey,* vol. 3, pp. 47-59, 2012.

[6] A. Sanchez, J. V. Ruiz, M. A.B., A. S. Montemayor, J. Hernandez, and J. J. Pantrigo, "Differential Optical Flow Applied to Automatic Facial Expression Recognition," *Neurocomputing*, vol. 74, pp. 1272-1282, 2011.

[7] A. Psaltis, "Optical Flow for Dynamic Facial Expression Recognition," Master, Faculty of Science, Utrecht University 2013.

[8] A. Besinger, T. Sztynda, S. Lal, C. J. Duthoit, J. I. Agbinya, B. Jap, *et al*., "Optical Flow Based Analyses to Detect Emotion From Human Facial Image Data," *Expert Systems with Applications*, 2010. doi: 10.1016/j.eswa.2010.05.063.

[9] F. Zhang, Gao, Y., Bakos, & J.D., "Lucas-Kanade Optical Flow Estimation on the TI C66x Digital Signal Processor," *IEEE High Performance Extreme Computing Conference (HPEC)*, pp. 1-6, 2014.

[10] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," presented at the Proceedings of the 7th Joint Conference on Artificial Intelligence, Vancouver British Coloumbia Canada, 1981.

[11] D. Patel and S. Upadhyay, "Optical Flow Measurement Using Lucas Kanade Method," *International Journal of Computer Applications*, vol. 61, pp. 6-10, 2013.

[12] T. Kanade, J. Cohn, and Y. L. Tian, "Comprehensive Database for Facial Expression Analysis," *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 46-53, 2000.