Temporal Modeling of Head Movement during Facial Expression via Dual Pivot Tracking System - A Case of Disgust

Fakhrul Hazman Yusoff, Rahmita Wirza O.K. Rahmat, Md. Nasir Sulaiman, Mohamed Hatta Shaharom, Hariyati Shahrima Abdul Majid

Abstract**—** Temporal modeling of facial expression is very important in various fields which include expression recognition, realism in computer animation and behavioral study in psychological field. While various researches are currently being done to capture the movement of facial features for its temporal property, not much is being said about the head movement during the facial expression process. The absence of head movement description can lead to incomplete expression description especially in the case of disgust where head movement can be a vital element for its depiction. Therefore, this paper proposes a method to track the movement of the head by using a dual pivot head tracking system (DPHT). In proving its usefulness, the tracking system will then be applied to track the movement of subjects depicting disgust. A simple statistical two-tailed analysis and visual rendering comparison will be made with a system that uses only a single pivot to illustrate the practicality of using DPHT. Results show that better depictions of expression can be implemented if the movement of the head is incorporated.

Keywords**—**Face Expression Modeling, Computer Graphics, Face Tracking, Face Animation.

I. Introduction

Facial expression modeling has been the interest of many researches due to its applicability in various areas such as recognition, robotics as well as computer animation. Many researches based their works on FACS, a face description method proposed by Paul Ekman [5]. Ekman's FACS face expression description can be found in works such as [1, 7, 12, 13, 19, 20]. While these works focus on the action unit (AU), works in term of head movement, that is part of the action descriptor (AD) is still lacking. Ekman, in his original work as can be attested from the action descriptor, defines some head movement specifically head turn-left, head turn-right, head-up, head-down, head tilt-left, head tilt right, head forward and head back [5, 6]. The head movement should be complemented with the AUs in order for a better depiction of facial expression. The expression of disgust and surprise for example will usually be accompanied with some movement of the head which signal the intensity of the action.

Existing works do not have a complete description of facial expression. Currently as most works focus on the AU, the graphical rendering usually is confined to face component movement (mouth, eyes, brows etc.). This reduces the realism factor of the depiction as the head movement is absence. If the head moves during the manifestation of expression, assessment will be incomplete since the existing tracking system such as [12, 13] just ignore the head movement. Take the case of disgust, the neck may move backward to signal displeasure. Without quantifying this movement, we may not be able to graphically depict the displeasure completely. Besides, any head movement that is depicted may be perceived to be the movement of the body instead of the specific neck movement.

Based on the AD as proposed by Ekman, the head movement must be inclusive of face, head and neck section. Therefore this paper will use the general term head to refer to combination of face, head and neck. It is the objective of this paper to propose a mechanism known as Dual-pivot Head Tracking (DPHT) System to track the head in order to quantify its movement. DPHT will then be applied on subjects depicting disgust to illustrate its applicability towards better realism in facial expression depiction.

Facial Action Coding System (FACS) by Paul Ekman is being used by many researches in the field of facial expression. However, the head motion description is not being considered by many of the researches. Ekman, through his Manual of FACS identifies many action descriptors including head motion [6]. Unfortunately, the head motion is not given priority in many studies. Tian, Kanade and Cohn [19] define templates to detect and track specific AU related muscle. The group's work is based on frontal view and focused on the AU. As AD description is not included, the visualization will be limited only to face. Moreover the group's work in 3D visualization will be limited due to the absence of depth information. Our proposed work on the other hand attempts to quantify movement of the head using both frontal and profile views. This enables 3D visualization as enough information is available to construct 3D coordinate.

Pantic and Patras look at rule-based description of AU. Their first paper focuses on profile view [12] while the second

Manuscript received October 22, 2008.
F. H. Yusoff Author is with the Universiti Putra Malaysia, Serdang, Malaysia. (e-mail: fhazman1975@yahoo.com.my).
R. W. Rahmat Author, is with the Universiti Putra Malaysia, Serdang, Malaysia. (e-mail: rahmita@fsktm.upm.edu.my).
M. N Sulaiman is with the Universiti Putra Malaysia, Serdang, Malaysia. (e-mail: nasir@fsktm.upm.edu.my).
M. H. Shaharom is with the Cyberjaya University College of Medical Sciences, Cyberjaya, Malaysia. (e-mail: hatta@cybermed.edu.my).
H. S. Abdul is with the International Islamic University, Malaysia. (e-mail: shahrima@iiu.edu.my).

paper looks at frontal view [13]. Valstar and Pantic [20] is a continuation of rule-based frontal description of AU. While there are frontal and profile elements, these works focus on fiducial points mainly for describing AU. Furthermore, no attempt is being made to implement it in 3D environment. The proposed method attempts to incorporate head motion which will be obtained from frontal and profile view in order for a 3D representation to be built. Using this approach a more encompassing movement can be visualized.

Some other works use specialized equipment to track movement. Kapoor and Picard [9] for example work on head nod and shake detector. The group use infra-red device to detect the movement. Meanwhile, Cohn, Schmidt, Gross and Ekman [2] use EMG (Facial Electromygraphic) device to quantify selected facial muscle during expression. Usage of devices may be applicable in lab environment. Unfortunately, it is intrusive and may hinder subjects from producing good facial expression. It is the goal of this work to avoid intrusion by using image processing approach to detect feature points.

MPEG-4 specification adapted FACS into FAP [11]. However, head movement is not incorporated in the specification. This will reduce the realistic value of the FAP since some of the facial expression includes the head movement. Similarly in an attempt to describe the archetypal expression using FAP, Raouzaiou et. al. did not implement the head movement where the work is confined on facial animation only [15].

Dornaika and Ahlberg [3] include some discussion on rotational property inferring 3D situation during their attempt to infer facial expression using template. The work even discuss on the yaw motion that exist during expression. However the implementation is confined to facial component instead of a more encompassing head movement.

Reisenzein, Bordgen and Holtbernd [17] attempt to investigate disassociation between emotion and facial display (surprise). However, their work only quantifies the eye area (pupil, brow) movement for sign of surprise. Incorporating description on the head movement may give more information on the surprise response.

The detachment of head movement issue and facial expression is such that various computer animation books do not touch on head movement while discussing facial expression. Face Animation and Computer Graphic books such as [8, 10, 16] do not touch on neck motion while discussing facial expression. The neck is actually an integral part of facial expression as pointed out by Roberts in [18] where he elaborates on the head positioning during facial expression.

It is apparent that a proper description of the neck together with the head movement needs to be introduced so that the movement can be better described. This paper proposes a 3D dual-pivot head movement approach (DPHT) to account for head turning and neck movement during facial expression.

II. Methods

Tracking the head movement which includes face and neck can be done by tracking some specific fiducial points. The technique of following fiducial points can be seen from the work of Pantic and Pantras in [12]. We extend the work of Pantic and Pantras so that the tracking of the neck section can be done.

Two pivot points and one secondary point are needed to track the movement of the head with respect to the neck. The first pivot (Point C) will be placed at the ear hole similar to fiducial point P15 of [12] while the second pivot (Point D) will be placed at the middle of the neck base. To enable tracking of the yaw, pitch and roll, another secondary point (Point P) which will be connected to Point C has to be tracked. We proposed the usage of nose tip which location is similar to fiducial point P4 of [12]. In other word, the fiducial points in our work are almost similar to [12] however our difference is the introduction of Point D. Point D is introduced as one of the points to be tracked to account for global movement as well as it enables the detection of neck movement.

To account for roll movement, extra points are needed to track the tilting of head. The nose holes areas can be tracked to achieve this objective. While [12] offers P5 for edge of nose hole, the proposed method will be using both sides (left nose and right nose) of P5 where label Point E and F will be given for each of the nose hole. The pivot points and related points are best described using the following Fig. 1 and Fig. 2:

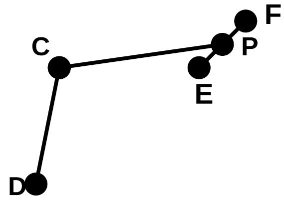


Fig. 1 DPHT Skeleton - Points to be tracked

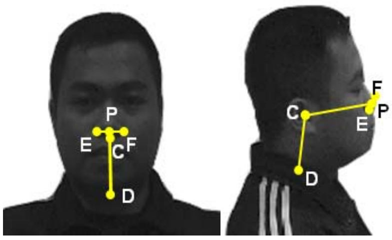


Fig. 2 Point Placement on the human head

Tracking Point D will be a bit of a challenge since we may not have distinct area for tracking. However, since Point D is just the base of DPHT and its main purpose is to play the role of tracking the global movement, the upper body part can be treated as one unit and will be represented as Point D. The head which includes the neck will then be tied to Point D via the linkage of Point D and C.

Assuming that scaling between frontal and profile images are equals, translation are equals, human face are symmetrical, and the frontal and profile view are orthogonal, establishing the 3D coordinate of Point P, C, D, E and F can be done using information obtained from frontal and profile views as follows:

$$P_x = \frac{E_{x:Frontal} + F_{x:Frontal}}{2} \qquad P_y = P_{y:Profile} \qquad P_z = P_{z:Profile} \tag{1}$$

$$D_x = D_{xFrontal} \qquad D_y = D_{y:Profile} \qquad D_z = D_{z:Profile} \tag{2}$$

$$C_x = D_x \qquad C_y = C_{y:Profile} \qquad C_z = C_{z:Profile} \tag{3}$$

$$E_x = E_{x:Frontal} \qquad E_y = E_{y:Frontal} \qquad E_z = P_z \tag{4}$$

$$F_x = F_{x:Frontal} \qquad F_y = F_{y:Frontal} \qquad F_z = P_z \tag{5}$$

By using (1), (2), (3), (4) and (5), head movement can be quantified. The information can be integrated with other facial expression information to create a better temporal rendering of a person expressing an expression.

III. Experiments

To demonstrate the importance of DPHT in describing facial expression, an experiment that elicits disgust expression has been devised. Thirty one subjects participate in the experiment. The subjects are selected from a pool of first semester students majoring in Computer Science. A simple random sampling approach is being used to obtain 31 students. To further minimize variations only students of a single race is being used for this experiment. A narration is read to the subject and followed by display of disgusting images. The session is repeated for three times where during each session a different disgusting image is displayed. The images consists of a person throwing up, a non-flushed toilet filled with faeces and an elephant defecating. Students are specifically requested to react to the narration and images by manifesting disgust. While this approach cannot be said as eliciting natural expression, it is still valid if we assume that human facial expression is a form of social signal where human manifest expression voluntarily to convey a message [14]. In this particular experiment, it is for the subject to socially signal disgust through manifestation of his facial expression.

Two video cameras are being used to capture the expression. One camera is positioned at the front and another one at the profile view of the subject. The two cameras that are orthogonally arranged enable the recovering of 3D coordinate with the assumption that the human face is symmetrical. As the focus is the general movement of the head, high precision of the exact location will not be the priority.

Upon recording, the video clip will be split into still image and DPHT will be applied to the subject head to track the movement. The movement use simple normalized cross correlation based (NCC) template matching [4] to track the fiducial points. For situations where occlusion cannot be avoided, a marker is put on subject to enable tracking.

In studying the motion of head during disgust, the angle of the line between Point D and C will be evaluated. Assuming the movement is mainly pull back movement with minimal side ways movement of the neck, the angle of line DC is calculated as follows:

$$\angle \mathrm{DC} = Tan^{-1}\left(\frac{D_y - C_y}{D_z - C_z}\right) \tag{6}$$

By using (6), the angle (in radian) can be tracked and this can be used as the value that quantifies the neck movement. To factor out individual posturing, first frame angle value will be deducted from all the recorded value as the first frame value usually indicate the personal posturing of each individual person.

In order to prove that the head movement exists during the facial expression, a simple statistical two-tailed test is being conducted using the following hypothesis:

H_0: $\mu = 0$ *rad*
H_1: μ I 0 *rad*

The testing will evaluate whether the average head movement (μ) is more than 0 *rad*. A significant head movement will indicate that during the facial expression of disgust, the neck movement is distinct and should be account as part of facial expression. A level of significant of α=0.05 is being used during the analysis to reject or accept H_1 .

To further emphasize the need for neck quantification, a graphical representation of head movement is rendered. The first representation illustrates the movement of head that incorporate the neck section as being proposed by this paper while the second representation uses fiducial points tracked using only points from the face area as proposed in work of Pantic and Pantras [12].

IV. Results

Based on the experiment of recording neck movement during expression of disgust for 31 subjects, the resulting graph will be as the following Fig. 3. Looking at Fig. 3, apparently the neck does not sit still during the manifestation of disgust. In analyzing the movement using two-tailed test as mentioned earlier, the z-score of the average movement for the samples of 31 recorded a value of -6.1, a value that far exceeds the 3 standard deviation of the normalized form. This means that H_0 can be rejected and the average head movement definitely is not 0 *rad*.

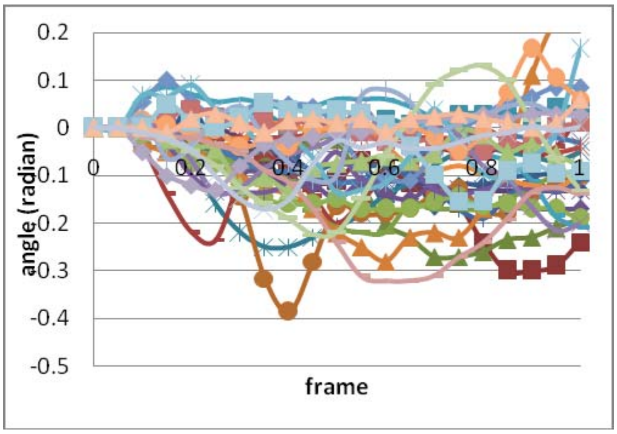


Fig. 3 Neck Movement for S1 to S31.

A closer look at the graph reveals that many samples recorded a trough like pattern. This is consistent with the expression of disgust where a person will pull back his head causing the neck to be pulled back together with the head. Fig. 4 shows a backward movement of some of the samples for clearer illustrations. While the trough varies from one person to another, the trough existence is consistent in almost all of the samples taken.

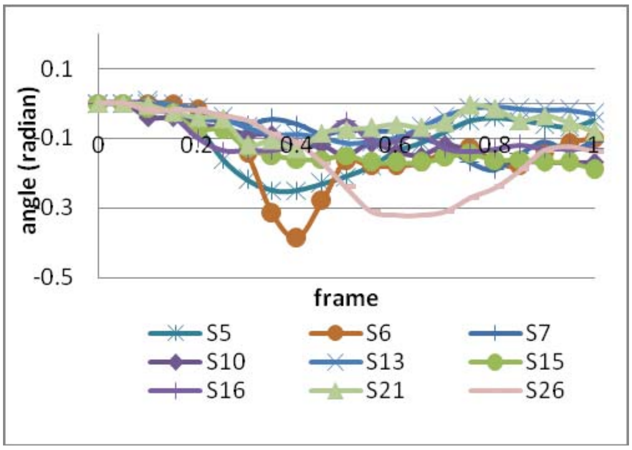
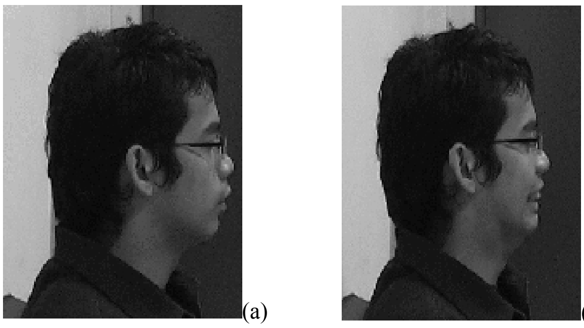


Fig. 4 Neck Movement for some of the samples.

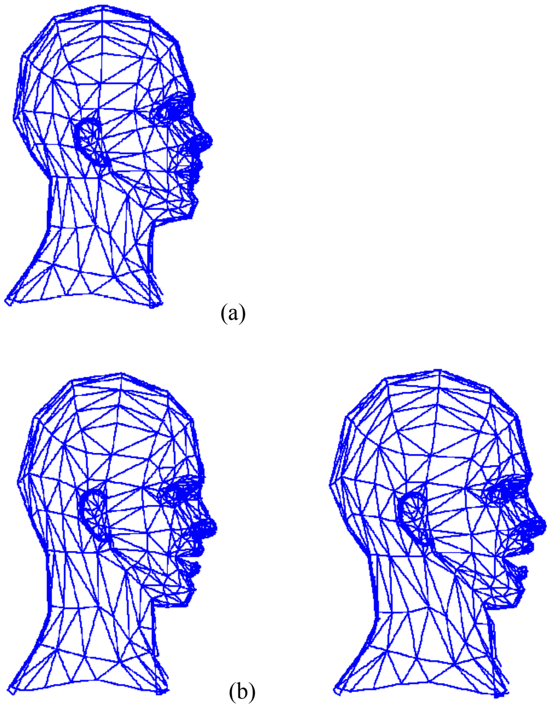
Assuming that the movement of roll, pitch and yaw is as minimal as possible, displaying disgust results in the head to be slightly pulled back. This causes the distance between the chin and the neck to be reduced as can be seen in Fig. 5(a) and Fig. 5(b).



(a) (b)

Fig. 5 (a) Initial Head Position (b) Face Depicting Disgust

The graphical representation of this reaction can be made possible by usage of DPHT where neck displacement information is available for visualization. Fig. 6 shows the different situation of the head movement. The first figure Fig. 6(a) depicts the initial condition of the head prior to any expression. Fig. 6(b) displays the expression manifestation using only fiducial points as proposed by Pantic and Pantras in [12]. Fig. 6(c) on the other hand illustrates the disgust manifestation using DPHT system.

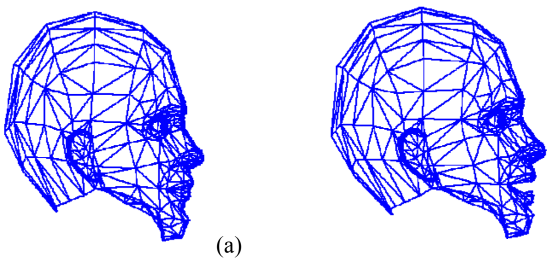


(a)

(b) (c)

Fig. 6 (a) Initial Position (b) Depicting Disgust – Fiducial Points do not include the neck area. (c) Depicting Disgust (DPHT) – Fiducial points includes the neck area.

Comparing Fig. 6(c) which is based on DPHT and Fig. 6(b), it is apparent, without the neck movement information, the expression appears to be less expressive. This is due to the movement mainly comes from the face section only. Fig. 6(c) carries a more intense manifestation with the neck leaning to the back. While the movement is small, we can clearly see it if compared to Fig. 6(b).



(a) (b)

Fig. 7 (a) Initial Position without Neck (b) Depicting Disgust without neck

Fig. 7(a) and (b) show the neutral and expression state without the usage of neck. As oppose to what can be illustrated in Fig. 6(c), the depiction of expression in Fig. 7(b) lacks manifestation of intensity and above all ignores the role being played by the neck during expressing an expression such as disgust. This example is yet another reason for neck to be treated as an integral element in facial expression study as it carries some weight in showing the intensity of expression.

V. CONCLUSION

Based from the experiments results and visualization, it is apparent that the neck movement contributes greatly towards understanding facial expression especially during expression such as disgust. Statistical analysis shows that the neck does not remain still during face expression of disgust. The neck exhibits significant movements which possibly can indicate the intensity of the expression. Therefore any face expression analysis such as Pantic and Pantras [12] will be more complete if neck movement analysis can be included with other facial expression descriptions. Furthermore by integrating the neck as part of head tracking system for facial expression, a more complete movement description can be obtained. This information enables more realistic computer graphics rendering of facial expression movements. Usage of two pivot points in DPHT specifically at the base of the neck and ear hole will make the temporal modeling to be smoother because apart from pitch, yaw and roll, additional movement such as pullback and push forward that are associated with the neck movement will also be defined. More importantly, the description of the additional movement will make the head movement definition to be more complete and accordance to the action descriptor (AD) as proposed by Paul Ekman in his FACS [5]. Last but not least, the proposed system can also be expanded to track the sideway movement of the head of which the best example can be seen from Indian classical dance where pitch, roll and yaw by themselves are not useful enough to describe the head movement. DPHT however must be implemented with care. The experiment above is carried out by focusing on limited movement of pullback of head thus the main indicator will be the movement of neck through displacement of Point C and D. A complete profiling on the other hand must also consider other standard movement such as yaw, pitch and roll. While not illustrated here due to complexity factor (and longer paper), the movement of yaw, pitch and roll can be easily calculated by tracking all the points of the DPHT namely Point C, D, E, F and P.

REFERENCES

[1] M. A. Bartlett, J. C. Hager, P. Ekman and T. Sejnowski, "Measuring facial expressions by computer image analysis," *International Journal of Psychophysiology*, Vol. 36(2), pp. 253-263, March 1999.

[2] J. F. Cohn, K. Schmidt, R. Gross, P. Ekman, "Individual Differences in Facial Expression: Stability over Time, Relation to Self-Reported Emotion and Ability to Inform Person Identification," *in IEEE International Conference on Multimodal Interfaces (ICMI 2002) 4*, pp. 491-496, Pittsburgh, PA, 2002.

[3] F. Dornaika, J. Ahlberg, "Face and Facial Feature Tracking Using Deformable Models," *International Journal of Image and Graphics (IJIG)*, Vol. 4(3), pp. 499 - 532, 2004.

[4] N. Efford, *Digital Image Processing - A Practical Introduction using Java*, Addison Wesley, Essex, England, 2000.

[5] P. Ekman, W. V. Friesen, *Facial Action Coding System - A Technique for the Measurement of Facial Movement*, Palo Alto California: Consulting Psychologist Press, 1978.

[6] P. Ekman, W. V. Friesen, C. Joseph, *Facial Action Coding System - The Manual on CD ROM*, Research Nexus Division of Network Information Research Corporation. Utah, USA, 2002.

[7] I. A. Essa, A. P. Pentland, "Facial expression recognition using a dynamic model and motion energy," in *Proceedings of the Fifth International Conference on Computer Vision*, pp. 360-367, USA, 1995.

[8] A. Jones, S. Bonney, *3D Studio Max 3 - Professional Animation*, New Riders, Indiana, USA, 2000.

[9] A. Kapoor, R. W. Picard, "A real-time head nod and shake detector," in *Proceedings of the 2001 workshop on Perceptive user interfaces*, pp. 1-5, Orlando, Florida, USA, 2001.

[10] J. Osipa, *Stop Staring - Face Modeling and Animation Done Right*, Sybex, Wiley Publishing, Indiana, USA, 2007.

[11] I. S. Pandzic, R. Forenheimer (ed.), *MPEG-4 Facial Animation - The Standard, Implementation and Application*, John Wiley & Sons. England, 2002.

[12] M. Pantic, I. Patras, "Temporal modeling of facial actions from face profile image sequences," in *IEEE International Conference on Multimedia and Expo*, pp. 49- 52, 2004.

[13] M. Pantic, I. Patras, "Detecting facial actions and their temporal segments in nearly frontal-view face image sequences," in *IEEE International Conference on Systems, Man and Cybernetics*, pp. 3358 - 3363, 2005.

[14] B. Parkinson, *Ideas and Realities of Emotion*, Routledge, London, U.K, 1995.

[15] A. Raouzaiou, N. Tsapatsoulis, K. Karpouzis and S. Kollias, "Parameterized Facial Expression Synthesis Based on MPEG-4 EURASIP," *Journal on Applied Signal Processing*, Vol. 2002 (10), pp. 1021-1038, 2002.

[16] P. Ratner, *3-D Human modeling and animation*, John Wiley & Sons. New Jersey, USA, 2003.

[17] R. Reisenzein, S. Bördgen, T. Holtbernd , D. Matz, "Evidence for strong dissociation between emotion and facial displays: the case of surprise," *Journal of Personality and Social Psychology*, Vol. 91(2), pp. 295-315. American Psychological Association, United States, 2006.

[18] S. Roberts, *Character Animation - 2D Skills for Better 3D*, Focal Press. Massachusetts, USA, 2004.

[19] Y. Tian, T. Kanade, J. F. Cohn, "Recognizing AU fro Facial Expression Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23(2), pp. 97-115, 2001.

[20] M. Vastar, M., Pantic, "Fully Automatic Facial Action Unit Detection and Temporal Analysis," in *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition Workshop*, pp. 149-156, 2006.