Evaluation of head pose features for stress detection and classification

Giorgos Giannakakis, *Member, IEEE*, Dimitris Manousos, Vaggelis Chaniotakis, Manolis Tsiknakis, *Member, IEEE*

Abstract— **This paper investigates variations in head pose features in response to specific stressors. A proper experiment consisting of neutral and stressful states was performed aiming to cover different types of stress affect. Then, features related to head movements and pose were computationally estimated and analyzed. Towards this direction, facial landmarks were fitted using Active Appearance Models (AAM). Using the 2D AAM facial landmarks, a 3D head pose model was estimated revealing head inclinations. Results indicate that specific stress conditions increase head mobility and mobility velocity, in both translational and rotational features. Even though stress modulates head movements and velocities, the most prominent increases are presented during tasks that include participant's speech. The degree and the intensity of the interaction effect between speech and stress should be investigated in more detail. The analysis reports that specific head pose features can be significant stress indicators that could contribute among other facial cues in reliable stress recognition.**

I. Introduction

Among different markers referred in the literature, the spatiotemporal behaviour of the head (i.e. head motion and head pose patterns) contain valuable affective information as part of an individual's non-verbal communication. Its interpretation, known as kinesics, can reveal various aspects of human activity as well as the existence, the intensity or the manifestation of specific emotional states.

Head pose estimation provides a rich source of information that can be utilized in several computer vision applications in human machine interaction, face recognition, attention estimation, affective computing, etc. Three-dimensional (3D) head pose estimation can also be an intermediate parameter in order to estimate other parameters such as user's spatial visual field. Nevertheless, vision-based pose estimation has plenty of challenges due to various image-dependent factors, e.g., lighting changes, camera distortion and projective geometry. An issue of great interest is the accurate estimation of 3D rotations: yaw, roll, and pitch angles, from two-dimensional (2-D) image parameters. An extensive survey regarding the 3D vision-based head pose estimation is provided in [1, 2].

Among the various implemented techniques that measure head motion, the techniques based on computer vision have lately gained in popularity because of their contactless approach, effectiveness and non intrusiveness. An issue of concern that should be taken into consideration is the fact that head displacement may be due to other body functions, such as speech, physical activity or other physiological functions, e.g. heart activity. An interesting approach, is the measurement of head micro movements as part of a methodology for the extraction of the heart rate [3, 4].

Head motion patterns are often employed as markers in the context of emotion recognition [5-7]. According to the [7], head orientation toward the camera is least frequent during boredom experiences, whereas moving the head downward is most typical of disgust while moving the head backwards is a typical motion linked to joy or pride. Other studies use head motion in the context of the evaluation of driver's attention during the guide and are based on the estimation of head pose between adjacent views in subsequent video frames [8, 9]. Head pose features are also used as indicators of depression [10], and negative affect [5]. Head gestures are considered to be modulated during stress conditions. When someone is stressed, there is greater head mobility [11] and motion of the head is more frequent and rapid [12, 13]. In the present study, we focus in quantifying the effect of stress in specific head motion related features.

II. Methods

A. Head movements estimation

In this section, the computational pipelines measure the motion of a person's head from a 2D video recording is described. The first step in the process is the face detection procedure; which is accomplished in the present study through the Viola-Jones method [14]. Subsequently, within the facial region of interest (ROI), alignment of the facial shapes variations into a mean facial shape is performed using an active shape Model (ASM) and facial images are warped and normalized in order to build a texture model. Finally, the shape and texture models generate a combined model, the Active Appearance Model (AAM) [15]. Using an AAM model, facial landmarks describing various facial parts are estimated as shown in Fig. 1. We used five landmarks in the vicinity of the nose (presented as red landmarks in Fig. 1) in order to detect the movements in the 2 dimensions (y-z). The specific

* This study was partially supported by IKY scholarships programme and co-financed by the action entitled "Reinforcement of Postdoctoral Researchers", in the framework of the Operational Programme "Human Resources Development Program" of the NSRF 2014 – 2020.

G. Giannakakis is with the Institute of Computer Science (ICS), Foundation for Research and Technology Hellas (FORTH), Vassilika Vouton, 70013, Iraklio, Greece (e-mail: ggian@ics.forth.gr).

D. Manousos is with the Institute of Computer Science (ICS), Foundation for Research and Technology Hellas (FORTH), (mandim@ics.forth.gr).

V. Chaniotakis is with the Institute of Computer Science (ICS), Foundation for Research and Technology Hellas (FORTH) and with the Department of Informatics Engineering, Technological Educational Institute of Crete (e-mail: chanvag@ics.forth.gr).

M. Tsiknakis is with the Department of Informatics Engineering, Technological Educational Institute of Crete and the Institute of Computer Science (ICS), Foundation for Research and Technology Hellas (FORTH) (e-mail: tsiknaki@ie.teicrete.gr).

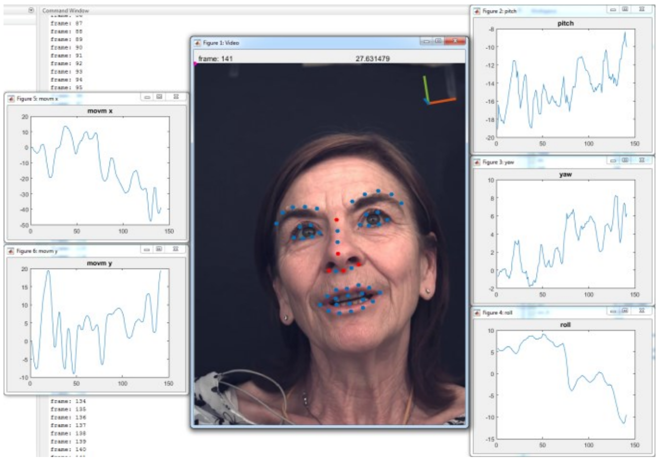

Figure 1. System's overview for given video recording. In the main window the video recording is presented with facial landmarks fitted with AAM and the pose axes estimation (pose estimation (blue vector: pitch, red vector: yaw, green vector: roll). The adjacent windows present the time series of the temporal displacements in y,z axis and the temporal evolution of pitch, yaw, roll angles.

landmarks were selected because they are characterized by the absence or the very low motion related to other factors, such as facial expressions, mouth movements, eye blinks, etc. In this way, head movements are isolated and measured more reliably.

The algorithm estimates the total head movement and head movements in terms of horizontal and vertical displacements in relation to predefined reference points. In addition, the variation of these movements form the total head velocity and corresponding velocities in the horizontal and vertical direction.

B. *Head pose estimation in 3D model*

Head pose was estimated by projecting a 3D face model onto 2D AAM. The projection is performed using the Pose from Orthography and Scaling with Iterations algorithm [16]. The algorithm requires known 2D position of at least four non-coplanar object's points, which in this study used 49 points derived from AAM. For the 3D model, a 3D face statistical anthropometric model was used. The algorithm estimates head pitch, yaw and roll displacement measured in angles in each frame forming the time series of these angles. Besides, the variation of these velocities provide the corresponding angular velocities.

TABLE I. EXPERIMENTAL TASKS, PHASES AND CONDITIONS EMPLOYED IN THE STUDY.

	Experimental phase	Affective State
Social Exposure		
1.1	Neutral (reference)	N
1.2	Self-describing	S
Emotional recall		
2.1	Neutral (reference)	N
2.2	Recall anxious event	S
2.3	Recall stressful event	S
Stressful images/Mental task		
3.1	Neutral/Stressful images	S
3.2	Mental task (Stroop Colour Word Task)	S
Stressful videos		
4.1	Neutral (reference)	N
4.2	Relaxing video	R
4.3	Adventure video	S
4.4	Psychological pressure video	S

Note: Intended affective state (N: neutral, S: stress, R: relaxed).

C. *Experimental procedure*

For the study reported in the present manuscript, the experimental protocol was established allowing the recording of participants' both neutral and stressed states, as shown in TABLE I. Different stressors were used categorized in phases aiming to cover different stress types.

Video acquisition was performed using a high performance color camera. In order to achieve the desired video quality, 2 high luminosity spotlights were placed at an angle behind the camera in order to produce ambient lighting conditions avoiding reflections and scattering effects. Each of the study participants were seated in front of a computer monitor. The camera was placed on a tripod at the back top of the monitor and at a distance of 90 cm with its field of view covering participant's face and possible movements during the experiment. At the beginning of the procedure, participants were informed about the whole procedure as well as about the terms of anxiety and stress.

The population of this study were 24 participants (7 women, 17 men) with age 47.3±9.3 years. The study was approved by the North-West Tuscany ESTAV (Regional Health Service, Agency for the Technical-Administrative Services of Wide Area) Ethical Committee. Data were recorded during the second data acquisition campaign (Semeoticons Reference Dataset SRD'15) within the context of the research project focused on the development of a computational platform monitoring cardio-metabolic risk [17]. From this dataset, the recordings regarding stress assessment (Semeoticons Reference Dataset for Stress Assessment SRDSA'15) were used.

D. *Feature extraction*

Once the 3D fitting was performed and the head pose was estimated, the time series of head movements, i.e. movement in y, z axis and the angles yaw, pitch, roll are extracted as shown in Fig. 2. Then, head movements and pose features were extracted from the displacement and rotational angles

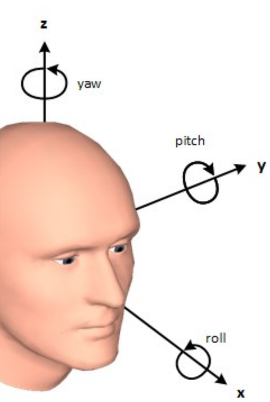

Figure 2. Illustration of head pose features: the translational (horizontal, vertical, lateral) and rotational (yaw, pitch, roll).

TABLE II. FEATURE EVALUATION ALONG EXPERIMENTAL TASKS PRESENTING VARIATION DURING STRESS CONDITION AND SIGNIFICANT DIFFERENCES

Feature	**Interview**		**Emotional Recall 1**		**Emotional Recall 2**		**IAPS images**		**Stroop Colour Word Task**	
	variation	**α**	**variation**	**α**	**variation**	**α**	**variation**	**α**	**variation**	**α**
total movement	↑	0.01		ns	↑	0.05	↑	0.01	↑	0.01
movement_y		ns		ns		ns		ns	↓	0.01
movement_z		ns		ns		ns	↑	0.01		ns
total velocity	↑	0.01		ns	↑	0.01	↑	0.05	↑	0.01
velocity_y	↑	0.01		ns	↑	0.05	↑	0.05	↑	0.01
velocity_z	↑	0.01		ns	↑	0.01		ns	↑	0.01
pitch		ns	↑	0.01	↑	0.01	↑	0.01	↑	0.05
yaw		ns	↓	0.01		ns		ns		ns
roll		ns		ns		ns		ns		ns
pitch velocity	↑	0.01		ns	↑	0.01		ns	↑	0.01
yaw velocity	↑	0.01		ns	↑	0.01	↑	0.01	↑	0.01
roll velocity	↑	0.01		ns	↑	0.05	↑	0.01	↑	0.01

Note: Features (α: confidence level, ns: non significant, ↓/↑ significant increase/decrease)

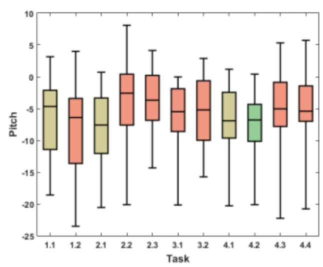

Figure 3. Variation of pitch across experimental tasks. Brown, red and green colours denote neutral, stressful and relaxed affective state respectively. (colour coding of affect state: brown = neutral, red = stress, green = relaxed)

time series (shown in Fig. 1). Specifically, the features: total movement, horizontal movement, vertical movement, yaw, roll, pitch. Besides, the velocities of these features were estimated: total velocity, velocity_y, velocity_z, pitch velocity, yaw velocity, roll velocity. A total of 12 features were estimated.

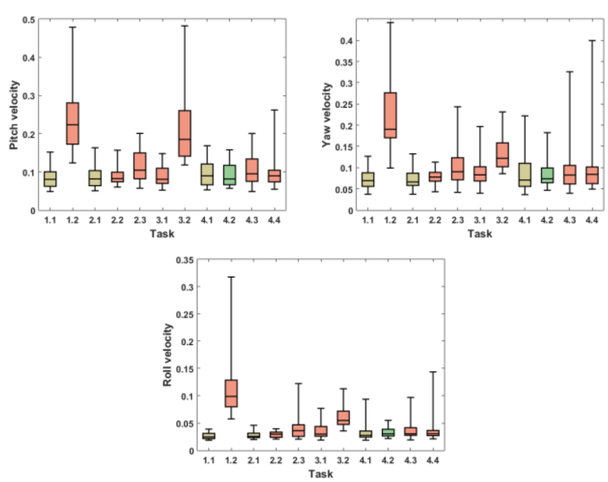

Figure 4. Pitch, yaw and roll velocities variation across tasks. There are significant increases during interview (1.2) and Stroop Colour Word Task (3.2) tasks related to speech

III. RESULTS

The head pose features were estimated using methods described in section II. Data were grouped by emotional state (2 levels: stress, no stress).

A. Feature evaluation

The head pose features were evaluated for significant differences. The overview of the results obtained, i.e. their variation across different stressors and their significance are presented in TABLE II.

It should be noted that under visual inspection time series of movements in y and z axis are highly correlated with pitch, yaw and roll. A one-way ANOVA was performed in order to examine the effects of stress on the head pose features in each experimental task. Comparisons among tasks were performed using as reference a neutral or relaxing recording at the beginning of each phase.

The pitch feature presented the most differences between emotional states (neutral, stress). The variation of pitch across experimental tasks is shown in Fig. 3. Specifically, pitch was significantly increased during the stressors of anxious event recall, stressful event recall, stressful images from IAPS, Stroop Colour Word Test. Regarding videos stimuli phase, pitch differences are achieved only when using as reference the relaxing video and not the neutral one. Increased pitch was typical across all stressful tasks meaning that head turns up during stress conditions. The yaw and roll feature differentiated the two emotional states only during anxious event recall and stressful video respectively.

B. Effect of speech

The variation of pitch, yaw and roll velocity features across experimental tasks are shown in Fig. 4. It can be observed that there are significant increases during the interview task (1.2) and the Stroop Colour Word Task (3.2) which both contain participants' speech. This finding is in accordance with the results of our previous study [18] in which the first dataset (Semeoticons Reference Dataset for Stress Assessment [SRDSA'14]) was analyzed.

TABLE III. SYSTEM'S BEST ACCURACIES CATEGORIZED FOR CLASSIFICATION SCHEME AND EXPERIMENTAL PHASE

Phase	Classification scheme		
	K-nn (k=3) (%)	**GLR (%)**	**SVM SVM (%)**
Social Exposure (interview)	98.6	97.9	97.2
Emotional Recall	82.99	68.2	85.42
Stressful images/ Stroop Color Word Task	85.42	76.25	80.0
Stressful videos	85.83	73.54	75.62

C. Classification

The effectiveness of feature set on stress recognition was assessed using feature selection methods and then feeding a classification scheme. A feature selection process was the first step in this procedure using sequential forward selection(SFS) [19]. The objective function for the selection method was the area under the ROC curve (AUC) [20] using the classifier under investigation classifying 2 emotional states. Then, the feature subset was feeded to the classification scheme, where the K-nearest neighbours (K-nn), Generalized Likelihood Ratio (GLR), Support Vector Machines (SVM) were tested. To evaluate the system's performance, the accuracy measure *((TP+TN)/(TP+TN+FP+FN))* was evaluated. Classifiers were tested using 10-fold cross-validation. The classification procedure was repeated 5 times. The average classification accuracy results of the social exposure and the emotion recall phases are presented in TABLE III.

It can be observed that during social exposure which includes the interview task high accuracy results are obtained. This can be attributed to the speech existence during this phase. Head movements should be decoupled from speech which is aimed to be performed in future study. The features that are selected are either pitch velocity, either roll velocity, either yaw velocity. As described in section III.B,TABLE II. head pose velocities are modulated by both stress and speech.

IV. DISCUSSION

The relevant literature of stress detection through head movements is limited whilst most studies try to correlate head movements with basic emotions. This paper reports investigation results of the head motion patterns during stress conditions. For this purpose, in the context of a relevant research project, two datasets (SRDSA '14 and SRDSA '15) of 23 and 24 participants respectively were generated with a year distance in between. Approximately 50% of the subjects were common in these two datasets. This study refers to the head pose features evaluation and analysis of the second dataset (SRDSA '15).

The results obtained indicate that during stress conditions, the head exhibits increased mobility with specific variations according to the stressors used. In fact, stress results in rapid movements which are revealed more prominently through velocity features. It should be noted that the most significant changes in head pose velocities are observed in tasks that include speech. A future study that also includes speech features would potentially allow us to investigate this relationship in more depth. A classification scheme was applied to assess the system's performance leading to fair stress recognition accuracies. Thus, head pose features may provide useful information about perceived stress states.

REFERENCES

[1] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation in computer vision: A survey," *IEEE Trans Pattern Anal Mach Intell,* vol. 31, pp. 607-626, 2009.

[2] J.-B. Kim, H.-I. Kim, and R.-H. Park, "Head pose estimation using a coplanar face model for human computer interaction," in *Consumer Electronics (ICCE), 2014 IEEE International Conference on*, 2014, pp. 560-561.

[3] R. Irani, K. Nasrollahi, and T. B. Moeslund, "Improved pulse detection from head motions using DCT," in *Computer Vision Theory and Applications (VISAPP), 2014 International Conference on*, 2014, pp. 118-124.

[4] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3430-3437.

[5] Z. Hammal, J. F. Cohn, C. Heike, and M. L. Speltz, "What can head and facial movements convey about positive and negative affect?," in *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on*, 2015, pp. 281-287.

[6] A. Adams, M. Mahmoud, T. Baltrusaitis, and P. Robinson, "Decoupling facial expressions and head motions in complex emotions," in *International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2015, pp. 274-280.

[7] H. G. Wallbott, "Bodily expression of emotion," *European journal of social psychology,* vol. 28, pp. 879-896, 1998.

[8] K. Liu, Y. Luo, T. Gyomei, and S. Yang, "Attention recognition of drivers based on head pose estimation," in *Vehicle Power and Propulsion Conference, 2008. VPPC'08. IEEE*, 2008, pp. 1-5.

[9] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness," *IEEE Transactions on intelligent transportation systems,* vol. 11, pp. 300-311, 2010.

[10] S. Alghowinem, R. Goecke, M. Wagner, G. Parkerx, and M. Breakspear, "Head pose and movement analysis as an indicator of depression," in *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, 2013, pp. 283-288.

[11] U. Hadar, T. Steiner, E. Grant, and F. C. Rose, "Head movement correlates of juncture and stress at sentence level," *Language and Speech,* vol. 26, pp. 117-129, 1983.

[12] W. Liao, W. Zhang, Z. I. Zhu, and Q. Ji, "A Real-Time Human Stress Monitoring System using dynamic bayesian network," NY, 2005.

[13] D. F. Dinges, R. L. Rider, J. Dorrian, E. L. McGlinchey, N. L. Rogers, Z. Cizman*, et al.*, "Optical computer recognition of facial expressions associated with stress induced by performance demands," *Aviation, space, and environmental medicine,* vol. 76, pp. 172-182, 2005.

[14] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision,* vol. 57, pp. 137-154, 2004.

[15] I. Matthews and S. Baker, "Active appearance models revisited," *International Journal of Computer Vision,* vol. 60, pp. 135-164, 2004.

[16] D. F. Dementhon and L. S. Davis, "Model-based object pose in 25 lines of code," *International journal of computer vision,* vol. 15, pp. 123-141, 1995.

[17] Y. Andreu, F. Chiarugi, S. Colantonio, G. Giannakakis, D. Giorgi, P. Henriquez*, et al.*, "Wize Mirror - a smart, multisensory cardio-metabolic risk monitoring system," *Computer Vision and Image Understanding,* vol. 148, pp. 3-22, Jul 2016.

[18] G. Giannakakis, M. Pediaditis, D. Manousos, E. Kazantzaki, F. Chiarugi, P. Simos*, et al.*, "Stress and anxiety detection using facial cues from videos," *Biomedical Signal Processing and Control,* vol. 31, pp. 89-101, 2017.

[19] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters,* vol. 15, pp. 1119-1125, 1994.

[20] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters,* vol. 27, pp. 861-874, 2006.