ON THE ROLE OF HEAD MOTION IN AFFECTIVE EXPRESSION

Atanu Samanta, Tanaya Guha

Electrical Engineering, Indian Institute of Technology, Kanpur, India

ABSTRACT

Non-verbal behavioral cues, such as head movement, play a significant role in human communication and affective expression. Although facial expression and gestures have been extensively studied in the context of emotion understanding, the head motion (which accompany both) is relatively less understood. This paper studies the significance of head movement in adult's affect communication using videos from movies. These videos are taken from the Acted Facial Expression in the Wild (AFEW) database and are labeled with seven basic emotion categories: *anger, disgust, fear, joy, neutral, sadness,* and *surprise*. Considering human head as a rigid body, we estimate the head pose at each video frame in terms of the three Euler angles, and obtain a time-series representation of head motion. First, we investigate the importance of the energy of angular head motion dynamics (displacement, velocity and acceleration) in discriminating among emotions. Next, we analyze the temporal variation of head motion by fitting an autoregressive model to the head motion time series. We observe that head motion carries sufficient information to distinguish any emotion from the rest with high accuracy and this information is complementary to that of facial expression as it helps improve emotion recognition accuracy.

Index Terms— affect analysis, head motion, facial expression.

1. INTRODUCTION

Automatic recognition and analysis of human emotion from non-verbal behavioral cues is important in many applications involving human-human and human-computer interaction. Among the various non-verbal cues, facial expressions is studied most extensively [1, 2, 3, 4], partly due to its obvious importance in affective expression. Since the development of the facial action coding system (FACS) [1] for decoding static facial expressions, significant effort has been put towards developing automated techniques for FACS-based action unit (AU) detection [2, 5, 6], and generic vision-based expression recognition systems [4, 7]. A handful of work has also attempted to use body gestures to understand human emotion [8, 9].

Rigid head motion is another non-verbal behavioral cue that plays important role in human communication. A spontaneous affective behavior is often supported by certain dynamics of head movements along with facial expressions and speech. Head motion is relatively well studied in the context of speech synthesis, where the objective is to synthesize realistic head motion from speech features [10, 11, 12]. Interpersonal coordination of head motion has been studied in the context of interaction between distressed couples [13], and between a mother and her infant [14]. A psychological experiment [15] reported that participants are able to recognize vocalists' emotional intent with $70 - 80\%$ accuracy from *only* head movements. This clearly shows that head motion is important for communicating affect. However, computational or engineering work on studying head motion for emotion understanding is scarce. One work focused on predicting emotions in continuous dimension (valence, arousal, expectation, intensity, and power) from magnitude and direction of two-dimensional head motion and head gestures, such as nods and shakes [16]. In another work [17] authors analyzed spontaneous affect using head motion along with temperature changes in infra-red thermal video. A recent work [18] isolated the head motion from facial expressions to demonstrate that the emotional information in head motion is complementary to the facial expressions. Another recent work studied the significance of head motion in positive and negative emotions of infants [19]. This work [19] noted that the angular velocity and acceleration of the head along pitch, yaw, and roll (see Fig. 1) are significantly higher for negative emotions as compared to the positive emotions.

In this work, we study the significance of head motion in adults' emotional expressions using video data. Note that our goal is not to design the best performing emotion recognition system, rather to understand how much information does head motion *alone* contain about an expressed emotion. We use the Facial Expression in the Wild (AFEW) database [20] consisting of more than 1000 video clips labeled with the seven basic emotion categories (anger, disgust, fear, joy, neutral, sadness, and surprise). Unlike previous studies that consider only two to three emotions (positive/negative and neutral), we study head motion for all seven emotion classes which allows for more fine-grained analysis. Considering human head as a rigid body, we first estimate the 3D head pose in terms of the three Euler rotation angles (*pitch*, *yaw*, *roll*) at each video frame, and obtain a time-series representation of head motion. We analyze these data using both non-parametric and parametric methods. Our non-parametric method computes the root-mean-square (RMS) values of the angular displacement, velocity, and acceleration of the human head for pitch, yaw and roll. Statistical tests are performed to understand whether or not these RMS measurements are significantly different among various emotions. We use the autoregressive (AR) model (parametric method) to capture the time evolution of head motion. In order to investigate the discriminative ability of head motion, emotion classification is performed using the AR coefficients and RMS measurements. We also investigate whether head motion is complementary or redundant to the information that facial expression provides for emotion recognition. Our study shows that the head motion dynamics alone has significant information to distinguish among emotions, and carries additional information that can improve facial expression-based emotion recognition systems.

2. DATABASE AND HEAD POSE ESTIMATION

Database: For our study on head motion analysis, we consider the *Acted Facial Expression in the Wild (AFEW)* [20] database. This database is chosen because it is one of the largest publicly available databases that contains emotions as close to spontaneous emotions as possible. This database contains video clips collected from fifty-four movies and labeled with seven basic emotion categories i.e. *anger, disgust, fear, joy, neutral, sadness,* and *surprise*.

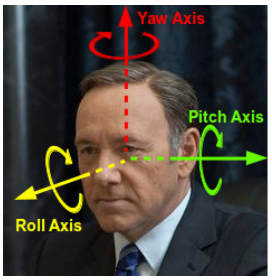


Fig. 1. Head pose defined in terms of the rotation of head about three principal axes - pitch, yaw and roll.

Table 1. Video clips used in our study for each emotion class.

Anger	Disgust	Fear	Joy	Neutral	Sadness	Surprise
161	107	103	192	180	147	105

The videos are of varying length and recorded at 25 frames/sec. The database is created through a semi-automatic approach, where the video clips are extracted via a recommender system based on subtitles and emotion labels are generated manually by human annotators. Each clip contains only one primary face or character.

Data preparation: The database (being 'in-the-wild') contains many videos where the apparent head movements are not entirely due to the persons' affective behavior, and instead, may be caused due to other factors, such as camera motion. In some videos only a part of the face is visible. In such situations, the estimated head pose (and subsequently the head motion) would be erroneous. For simplicity and to adhere to our original goal of studying head movement in emotion, we manually removed such videos from the database. Even after removal of these clips, each class contains more than 100 samples (see Table 1).

Head motion estimation: Considering human head as a rigid body, let us define 3D head pose or orientation at i^{th} frame as $\boldsymbol{\theta}^i = \left[\theta_p^i, \theta_y^i, \theta_r^i\right]^T$ where θ_p^i, θ_y^i, and θ_r^i are the three Euler angles, referred as *pitch*, *yaw*, and *roll* (see Fig.1). The head motion for a given video sequence is thus represented as a time-series $\boldsymbol{\Theta} = \left\{\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, ..., \boldsymbol{\theta}^N\right\}$ where N is number of frames in the video sequence. The 3D head pose $\boldsymbol{\theta}^i$ at a given frame (see Fig.2) is estimated using the *incremental face alignment* method [21]. This method uses a parameterized facial shape model which combines the non-rigid shape variation and the rigid transformation of the global shape. The parameters depend on the 3D head pose/orientation $(\theta_p, \theta_y, \theta_r)$, scale, and translation responsible for the rigid transformation required to align the face shape on the face image in the video frame. At the frames where face could not be detected, cubic interpolation is used to estimate the missing head pose. Gaussian smoothing is applied to remove small noise that may be present due to errors in estimation and interpolation.

3. HEAD MOTION ANALYSIS

The objective of this study is to understand how much information does head motion carry about an adult's expressed emotion. We *hypothesize* that humans have different head motion characteristics for different emotions. The first step towards validating this hypothesis is to obtain an effective representation of head motion activity. We employ two methods for characterizing head motion from its time series representation. One approach is to represent head motion in terms of the RMS values of the angular displacement, velocity and acceleration of pitch, yaw, and roll. This is a non-parametric approach and the RMS values are indicative of the overall head motion dynamics. In order to capture the temporal variation in head motion, we use an autoregressive (AR) model - a popular parametric method for time series modeling. To test our hypothesis, a set of statistical significance tests and classification experiments are performed.

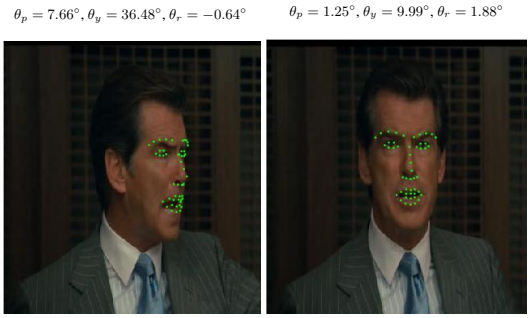


Fig. 2. Examples of detected facial landmark points and estimated head pose in terms of θ_p, θ_y, and θ_r (best viewed in color).

3.1. RMS measurements of head movement dynamics

We compute the *angular displacement*, $\boldsymbol{\Theta}_d = \left[\boldsymbol{\theta}_d^1, \boldsymbol{\theta}_d^2, ..., \boldsymbol{\theta}_d^N\right]$ by subtracting the mean head pose from the head pose at each video frame. Note that $\boldsymbol{\Theta}_d \in \mathbb{R}^{3\times N}$, where the three rows correspond to pitch, yaw and roll. Then the first and second derivatives of $\boldsymbol{\Theta}_d$ is computed to obtain the *angular velocity* $\boldsymbol{\Theta}_v = \left[\boldsymbol{\theta}_v^1, \boldsymbol{\theta}_v^2, ...\boldsymbol{\theta}_v^{N-1}\right]$, and the *angular acceleration* $\boldsymbol{\Theta}_a = \left[\boldsymbol{\theta}_a^1, \boldsymbol{\theta}_a^2, ..., \boldsymbol{\theta}_a^{N-2}\right]$.

$$\boldsymbol{\theta}_d^i = \boldsymbol{\theta}^i - \frac{1}{N}\sum_{j=1}^{N}\boldsymbol{\theta}^j \tag{1}$$

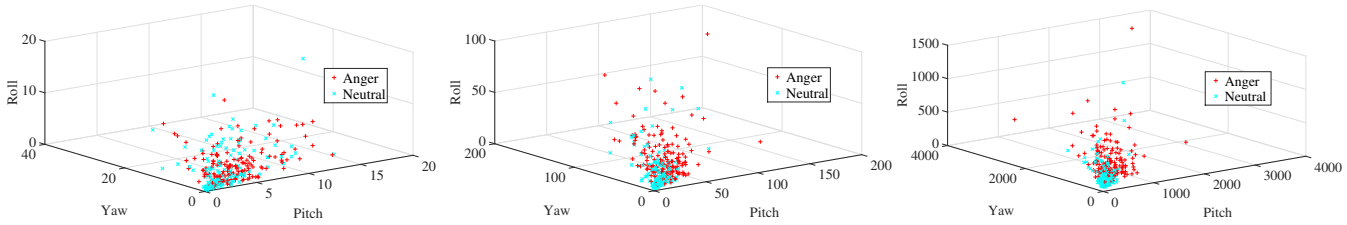
$$\boldsymbol{\theta}_v^i = \left(\boldsymbol{\theta}_d^{i+1} - \boldsymbol{\theta}_d^i\right) \times FrameRate \tag{2}$$

$$\boldsymbol{\theta}_a^i = \left(\boldsymbol{\theta}_v^{i+1} - \boldsymbol{\theta}_v^i\right) \times FrameRate \tag{3}$$

The RMS values of the nine time-series (angular displacement, velocity, and acceleration of pitch, yaw, and roll) are computed for every video, for each emotion class. Fig.3 shows the nine RMS values for all videos pertaining to anger and neutral. Qualitative observation of the plots in Fig.3 suggests that the RMS measurements of the head motion dynamics are quite different for the two emotions.

Statistical significance tests: To test our hypothesis that the head motion characteristics (measured in terms of the nine RMS values) of different emotions are significantly different we perform analysis of variance (ANOVA) followed by post-hoc multiple comparison and paired t-test. We perform ANOVA separately for each of the nine RMS measurements. The ANOVA results ($p < 0.05$, for all RMS measurements) indicate that RMS measurements of the head motion dynamics of at least one emotion category are significantly different from those of other emotion categories.

Following the results of ANOVA (suggesting significant difference in head dynamics across emotions), a post-hoc multiple comparison and paired t-tests are performed (see Table 2). The post-hoc multiple comparison (in Fig.4) provides a detail report of the emotions for which the measurements are significantly different. The RMS measurements of the emotions are considered significantly different (with 5% significance or with 95% confidence level) if their the confidence intervals are non-overlapping. Our observations are:



(a) Angular displacement ($^\circ$) (b) Angular velocity ($^\circ/sec$) (c) Angular acceleration ($^\circ/sec^2$)

Fig. 3. Scatter plots of RMS measurements of head motion dynamics for anger vs. neutral. (best viewed in color)

Table 2. Paired t-test results for one vs. all emotions. The reported numbers are p-values.

	Displacement			Velocity			Acceleration		
Emotion	Pitch	Yaw	Roll	Pitch	Yaw	Roll	Pitch	Yaw	Roll
Anger	**0.000**	0.432	0.851	**0.000**	**0.024**	**0.015**	**0.000**	**0.015**	**0.041**
Disgust	0.389	**0.000**	**0.002**	0.222	**0.031**	0.167	**0.001**	0.839	0.220
Fear	0.322	0.096	0.067	0.083	0.052	0.722	0.496	0.405	0.605
Joy	**0.007**	0.747	**0.001**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Neutral	**0.000**	**0.000**	**0.002**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Sadness	0.510	0.693	0.876	**0.000**	**0.002**	**0.001**	**0.000**	**0.000**	**0.000**
Surprise	**0.000**	0.414	**0.005**	**0.000**	0.310	**0.000**	**0.000**	**0.006**	**0.000**

(i) the differences among emotions are more significant in the angular velocity and acceleration, as compared to angular displacement, (ii) the emotions - anger, joy, and neutral - are highly distinguishable from other emotion classes in terms of almost all RMS measurements, (iii) the RMS measurements for sadness, surprise, and neutral are low and of similar values, (iv) the angular velocity and acceleration of *pitch* of the head are significantly higher for anger and joy as compared to the most of the emotions, (v) the angular velocity and acceleration of *roll* are significantly higher for joy than those for any of the six other basic emotions.

Classification experiments: We next investigate how discriminative the RMS measurements are in terms of emotion classification. A classification experiment was performed using the RMS measurements as features of head motion. We use a k-Nearest Neighbor (kNN) classifier ($k = 5$) where the head motion extracted from each video clip is represented as a 9-dimensional feature (corresponding to the 9 RMS measurements). Table 3 presents the results (10-fold cross validation) of classifying any *one emotion from the rest*, where the accuracy is around 80%. This results comply with the previously reported psychological study [15], where participants have recognized one of the three emotions with about $70 - 80\%$ accuracy. Next, we perform a multiclass classification with 7 classes. For this purpose, 80% of the data is used for training, and the rest for testing. The overall classification accuracy is about 34% (see Table 4), which is approximately double the accuracy of a random guess. Consistent with the significance test results, emotion classes, anger, joy, and neutral are recognized with a higher accuracy, and sadness, and surprise is often miss-classified as neutral.

3.2. Autoregressive modeling

AR models are popular for describing time-varying processes. We use an AR model to capture the temporal dynamics of the head motion data. We fit a third order auto-regressive model, AR(3), (Eq.4) to each of the three time-series corresponding to pitch, yaw and roll.

Table 3. Accuracy of one-against-all emotion classification using RMS measurements of head motion dynamics.

Anger	**Disgust**	**Fear**	**Joy**	**Neutral**	**Sadness**	**Surprise**
0.81	0.82	0.79	0.83	0.83	0.82	0.83

Table 4. Confusion matrix for emotion classification using RMS measurements of head motion dynamics.

	Anger	Disgust	Fear	Joy	Neutral	Sadness	Surprise
Anger	**0.47**	0.00	0.04	0.25	0.16	0.08	0.00
Disgust	0.16	**0.13**	0.06	0.09	0.25	0.31	0.00
Fear	0.17	0.07	**0.03**	0.23	0.43	0.07	0.00
Joy	0.25	0.05	0.05	**0.56**	0.05	0.04	0.00
Neutral	0.07	0.02	0.02	0.06	**0.61**	0.20	0.02
Sadness	0.20	0.05	0.07	0.07	0.38	**0.18**	0.05
Surprise	0.09	0.07	0.00	0.22	0.50	0.09	**0.03**

The order three is determined based on the Bayesian information criteria (BIC) computed using a set of randomly selected typical time series from each emotion class.

$$\theta(t) = a_0 + a_1\theta(t-1) + a_2\theta(t-2) + a_3\theta(t-3) \qquad (4)$$

Classification experiments: A multiclass emotion recognition experiment is performed (as before) using the AR(3) coefficients as features. The resulting average accuracy is 22% (lower than that obtained using RMS values). We suspect that this is due to the video clips in the database being too short, and not having the entire emotion profile (neural-emotion-neutral). From the confusion matrix in Table 5, we observe that anger, joy and neutral emotions have higher accuracy while sadness and surprise are confused with neutral and happy. This is consistent with the RMS-based results presented in Table 4.

3.3. Comparison with facial expression

So far, we have established that head motion itself contains significant information about a person's affect. However, it is not yet clear if the information extracted from head motion is already contained in facial expressions (more explicit affect information) or is complementary to facial expressions. To investigate this, we compare emotion recognition accuracy obtained using only facial expressions with that obtained after adding head motion to it. Facial expression dynamics are captured using local binary pattern - three orthogonal planes (LBP-TOP) features [7] - a popular feature for facial expression recognition [22, 23]. Before feature extraction, face alignment

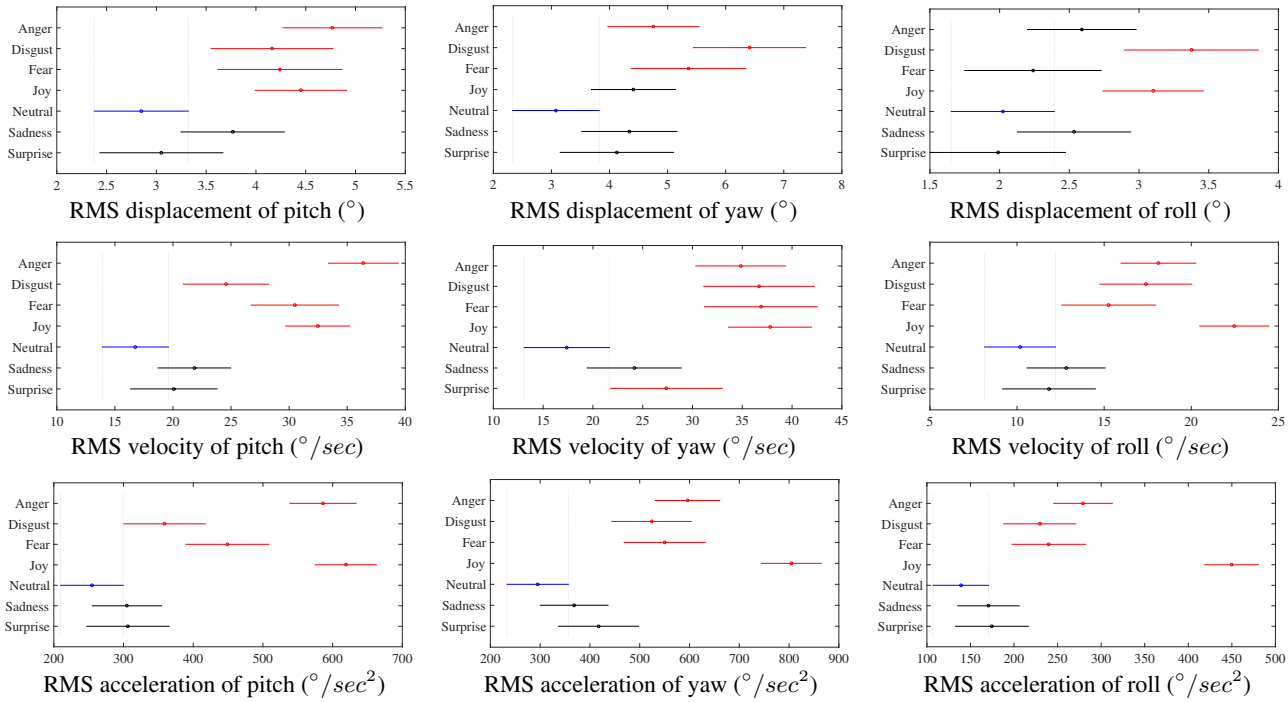


Fig. 4. 95% confidence intervals of in-class means of RMS measurements. The RMS measurements are significantly different for emotion categories for which the confidence intervals are non-overlapping. For example measurements for neutral emotion (blue) are significantly different from the measurements for emotions whose confidence intervals are displayed in red.

Table 5. Confusion matrix for emotion classification using AR(3) coefficients of head motion dynamics.

	Anger	Disgust	Fear	Joy	Neutral	Sadness	Surprise
Anger	**0.31**	0.10	0.06	0.28	0.19	0.06	0.00
Disgust	0.18	**0.09**	0.05	0.27	0.23	0.18	0.00
Fear	0.19	0.00	**0.10**	0.19	0.42	0.05	0.05
Joy	0.19	0.06	0.05	**0.30**	0.24	0.16	0.00
Neutral	0.17	0.08	0.03	0.22	**0.42**	0.08	0.00
Sadness	0.10	0.04	0.00	0.41	0.31	**0.14**	0.00
Surprise	0.25	0.15	0.00	0.25	0.35	0.00	**0.00**

is done using landmark points obtained by incremental face alignment [21]. For classification using both facial expression and head motion, the nine RMS measurements of the head motion dynamics are concatenated with the LBP-TOP features. The results are presented in Table 6. Clearly, the addition of head motion dynamics improves the emotion recognition performance of the facial-expression based system. Note that this observation is in compliance with the previous study [18]. To put the results into context, we also mention the baseline accuracy obtained for this database as part of an emotion recognition in the wild challenge [22].

4. CONCLUSION

The main contribution of this paper is to present a systematic study of analyzing the significance of head motion in conveying affect or emotion. Considering human head as a rigid body, head motion was represented as a time-series of head pose (defined in terms of pitch, yaw and roll), estimated at each frame of video sequence. We investigated the dynamics of head motion along pitch, yaw and roll for seven basic emotions. Through statistical tests and discriminative analysis (classification) we have shown that head motion *alone* carries adequate information to distinguish any basic emotion from the rest. More importantly, the information contained in head movements is noted to be complementary to that contained in facial expressions, since adding head motion information can significantly improve the accuracy of facial expression-based emotion classification system. Another observation is that head motion (angular velocity and acceleration) is significantly faster for *anger* and *joy*, while slower for *sadness*, *surprise*, and *neutral*. Due to very short video clips, we could not employ time-series modeling techniques to their full strength which could possibly generate further insights. In summary, this study has shown that head motion alone carries significant information about human emotion, and can inform the development of robust affect analysis system.

Table 6. Classification results for facial expressions and head motion

Non-verbal cue	Accuracy (in %)
Facial expression (LBP-TOP)	26.84
Head motion (RMS measures)	34.23
Facial expression + Head motion	**36.15**
Facial expression AFEW baseline [1] [22]	39.33

[1]This paper uses LBP-TOP with support vector machine (SVM), and a sophisticated face alignment method. This result is mentioned to give an idea of the standard classification accuracy on this database.

5. REFERENCES

[1] Paul Ekman and Wallace V Friesen, "Measuring facial movement," *Environmental psychology and nonverbal behavior*, vol. 1, no. 1, pp. 56–75, 1976.

[2] Y-I Tian, Takeo Kanade, and Jeffrey F Cohn, "Recognizing action units for facial expression analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 2, pp. 97–115, 2001.

[3] Zhihong Zeng, Maja Pantic, Glenn I Roisman, and Thomas S Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 1, pp. 39–58, 2009.

[4] Caifeng Shan, Shaogang Gong, and Peter W McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[5] Yan Tong, Wenhui Liao, and Qiang Ji, "Facial action unit recognition by exploiting their dynamic and semantic relationships," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 10, pp. 1683–1699, 2007.

[6] Michel F Valstar and Maja Pantic, "Fully automatic recognition of the temporal phases of facial actions," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 1, pp. 28–43, 2012.

[7] Guoying Zhao and Matti Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 6, pp. 915–928, 2007.

[8] Caifeng Shan, Shaogang Gong, and Peter W McOwan, "Beyond facial expressions: Learning human emotion from body gestures.," in *BMVC*, 2007, pp. 1–10.

[9] Hatice Gunes and Massimo Piccardi, "Bi-modal emotion recognition from expressive face and body gestures," *Journal of Network and Computer Applications*, vol. 30, no. 4, pp. 1334–1345, 2007.

[10] Hani C Yehia, Takaaki Kuratate, and Eric Vatikiotis-Bateson, "Linking facial animation, head motion and speech acoustics," *Journal of Phonetics*, vol. 30, no. 3, pp. 555–568, 2002.

[11] Carlos Busso, Zhigang Deng, Ulrich Neumann, and Shrikanth Narayanan, "Natural head motion synthesis driven by acoustic prosodic features," *Computer Animation and Virtual Worlds*, vol. 16, no. 3-4, pp. 283–290, 2005.

[12] Carlos Busso, Zhigang Deng, Michael Grimm, Ulrich Neumann, and Shrikanth Narayanan, "Rigid head motion in expressive speech animation: Analysis and synthesis," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1075–1086, 2007.

[13] Zakia Hammal, Jeffrey F Cohn, and David T George, "Interpersonal coordination of headmotion in distressed couples," *IEEE transactions on affective computing*, vol. 5, no. 2, pp. 155–167, 2014.

[14] Zakia Hammal, Jeffrey F Cohn, and Daniel S Messinger, "Head movement dynamics during play and perturbed mother-infant interaction," *IEEE transactions on affective computing*, vol. 6, no. 4, pp. 361–370, 2015.

[15] Steven R Livingstone and Caroline Palmer, "Head movements encode emotions during speech and song.," *Emotion*, vol. 16, no. 3, pp. 365, 2016.

[16] Hatice Gunes and Maja Pantic, "Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners," in *International conference on intelligent virtual agents*. Springer, 2010, pp. 371–377.

[17] Peng Liu and Lijun Yin, "Spontaneous facial expression analysis based on temperature changes and head motions," in *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*. IEEE, 2015, vol. 1, pp. 1–6.

[18] Andra Adams, Marwa Mahmoud, Tadas Baltrušaitis, and Peter Robinson, "Decoupling facial expressions and head motions in complex emotions," in *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on*. IEEE, 2015, pp. 274–280.

[19] Zakia Hammal, Jeffrey F Cohn, Carrie Heike, and Matthew 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*. IEEE, 2015, pp. 281–287.

[20] Abhinav Dhall, Roland Goecke, Simon Lucey, and Tom Gedeon, "Collecting large, richly annotated facial-expression databases from movies," *IEEE MultiMedia*, vol. 19, no. 3, pp. 34–41, 2012.

[21] Akshay Asthana, Stefanos Zafeiriou, Shiyang Cheng, and Maja Pantic, "Incremental face alignment in the wild," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014, pp. 1859–1866.

[22] Abhinav Dhall, OV Ramana Murthy, Roland Goecke, Jyoti Joshi, and Tom Gedeon, "Video and image based emotion recognition challenges in the wild: Emotiw 2015," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ACM, 2015, pp. 423–426.

[23] Guoying Zhao and Matti Pietikainen, "Experiments with facial expression recognition using spatiotemporal local binary patterns," in *IEEE International Conference on Multimedia and Expo*. IEEE, 2007, pp. 1091–1094.