

# Towards an Affect Sensitive Interactive Companion

Hamit Soyel, Peter W. McOwan

*School of Electronic Engineering and Computer Science  
Queen Mary University of London  
London E1 4NS, United Kingdom*

---

## Abstract

As robots are increasingly being viewed as social entities to be integrated in our daily lives, social perceptive abilities seem a necessary requirement for enabling more natural interaction with human users. In this paper, we present an interaction scenario where user play chess with an iCat robot and propose an affect recognition system that uses computational models to automatically extract visual features allowing the detection of the level of engagement with a social robot that acts as a game companion. Experimental results show that the multimodal integration of head direction information with facial expressions displayed by the user improves the recognition of the user's affective states.

*Keywords:* Action Units, Affective Computing, Human-Robot Interaction

---

## 1. Introduction

The design of an affect recognition module based on the interpretation of the user's behaviour is the first step towards the generation of a high quality Human-Robot-Interaction (HRI). Interactive companions are an example of

---

*Email address:* [hsoyel@eecs.qmul.ac.uk](mailto:hsoyel@eecs.qmul.ac.uk) (Hamit Soyel)

social embodiments which may benefit from the integration of such a “affect recognition module”. Interactive companions can be useful in many applications: they can be employed as personal assistants in smart environments, as interactive toys for therapy and rehabilitation purposes, they can provide additional functionalities to assist carers, healthcare workers, etc. For these companions to be able to cover these roles it is necessary that they are endowed with social capabilities and are sensitive to what happens in the external world, with a special attention to what the user feels or communicates. While significant advances have been made in the field of affect recognition over the past decade [1][2] the design of such a module to be integrated in a HRI framework has not been extensively addressed yet. The need for artificial companions to work in the user’s own social settings and to create long-term relationships with humans requires then research on affect recognition to be taken beyond the state of the art.

Many of the affect recognition systems described in the literature mainly focused on the recognition of basic emotions (e.g., joy, sadness, disgust, surprise, fear, anger, etc.) [3]. While the automated recognition of more complex states has started to receive some attention only lately [4], research on artificial companions requires the design of an affective framework in which the companion’s affect sensitivity goes beyond the ability to recognise prototypical emotions, and allows for more variegated affective signals conveying more subtle states such as, boredom, interest, frustration, agreement, etc., to be captured. It is important to stress that the inclusion of affect representation into a framework for affect recognition is of primary importance. Incorporating models and paradigms developed by psychologists for the clas-

sification of affective states [5] is a pressing need, but is still a challenging issue. Strengthening the connection with psychological models would allow for the first steps towards the detection of more complex affective states (e.g., appraisals, blends of emotions, preferences, mood, attitudes, etc.) to be undertaken.

The design of most existing affect recognition systems was largely based on databases of acted affective expressions [6]. While acted affective expressions, contrary to spontaneous expressions, can be defined precisely, allow for the recording of several affective expressions for the same individual, and can be characterised by very high quality, they often reflect stereotypes and exaggerated expressions, not genuine affective states, and they are often decontextualised [1]. The design of an artificial companion would certainly benefit from the development of affect detectors which are trained and tested with spontaneous, real-life expressions. Collection of facial behaviour in naturalistic contexts involves several issues, such as the difficulty of recording several emotional reactions for the same individual. Nevertheless, this is an issue that must be addressed in the design of an affect sensitive companion, in which personalisation plays an important role.

Another important issue for affect sensitive artificial companions is the need for a multimodal affect recognition system. It is expected that a companion is endowed with the ability to analyse different types of affective expressions, depending on the specific scenario of interaction with the user. Fusing different affective cues can allow for a better understanding of the affective message communicated by the user to be achieved. The development of novel methods for multimodal fusion should take into consideration

what are the underlying relationships and correlation between the feature sets in different modalities [1], how different affective expressions influence to each other and how much information each of them provides about the communicated affect.

Artificial companions have to be designed so as to be able to work in the users' own settings. This requires a companion's affect recognition system to be robust in real world conditions: face detectors and body and facial features tracking systems which are robust to occlusions, illumination changes, non-rigid head motions, etc., are some of the most important requirements for a companion to successfully work in real environments [7]. Real world scenarios means that the companion must infer the user's state in real-time. This poses several issues, such as, for example, the segmentation and the analysis of the temporal dynamics of face or body gestures and expressions, since the possibility for a user's affective state to start at any time is a crucial factor in real-time affect recognition [8][9].

An important issue to be considered in the design of an affect recognition system for artificial companions is represented by taking into consideration the context in which an affective expression is displayed (e.g., characteristics of the person expressing the emotion, environment in which the emotion is displayed, what the person is doing (i.e., their task), underlying mood, behaviour displayed by the companion, presence of other people, etc.). As suggested by [7], there can be as many emotions as the patterns of appraisal results. This highlights the importance of the evaluation of a stimulus event for the generation of the emotional response. In the same way, artificial companions must be able to evaluate how the recognised affective state relates



Figure 1: A user playing chess with the iCat.

with the conditions external to an individual that elicited the emotional response. In this paper we focus on a specific interaction scenario where user play chess with an iCat robot [10] as depicted in Figure 1 and we propose an affect recognition system that uses computational models to automatically extract contextual information and visual features allowing the detection of the level of engagement with a social robot that acts as a game companion. Built on a 3D distribution of facial features extracted by faceAPI we can track a face and recover the valence and interest towards the game companion to infer the player’s emotional state. The feeling index fuses data on the facial expression and head direction information to accurately and robustly predict the engagement of the player with the companion. Experimental results show that the multimodal integration of head direction information with facial expressions displayed by the user improves the recognition of the user’s affective states.

## 2. Modalities and Affect Recognition Framework

Contractions of facial muscles induced movements of the facial skin and changes in the appearance of facial features such as eyebrows, nose and mouth. Their shape and location can alter immensely with facial expressions and head pose variations. To be able to reason about the shown expression and the facial muscle actions that produce it, one must first detect the current appearance of the facial features. To do so, we track a set of facial points illustrated in Figure 2, the locations of which alter as the current appearance of the facial features changes with facial expressions and head pose variations. A standard webcam, positioned in front of the user captures the nonverbal behaviour displayed by the children during the game and the interaction with the robot. The system performs tracking of head movements and salient facial points via faceAPI, a real-time face tracking toolkit from Seeing Machines<sup>1</sup>, and estimates the gaze direction of the user based on head direction and rotation data. Furthermore, geometrical facial features extracted from the tracked facial points are used to detect users affective behaviour.

### 2.1. Parameters for AU Detection

Although muscle actions are of high importance one is unable to track them analytically without resorting to explicit sensors [11]. However, a subset of them can be deduced from their visual results, that is, the deformation of the facial tissue and the movement of some facial surface points. This reasoning resembles the way that humans visually perceive emotions, by noticing

---

<sup>1</sup><http://www.seeingmachines.com/product/faceapi>

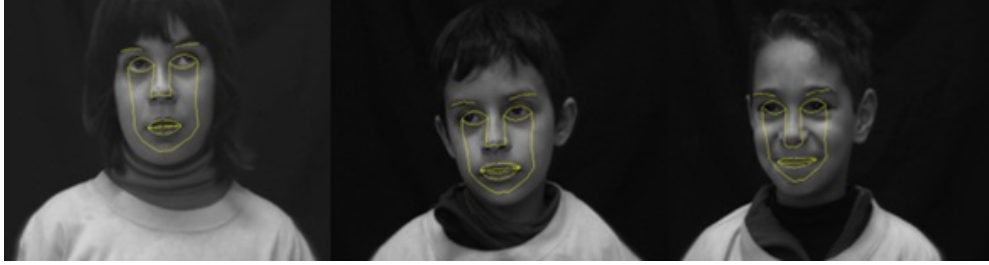


Figure 2: Facial feature tracking.

specific features in the most expressive areas of the face, the regions around the eyes and the mouth [12]. In our study, we are confined in these three components and then determine 3D distributions of the facial feature points which are representative of the boundary between these components and skin. By using the symmetry of the human face we have optimized the number of facial features used by 23 [13]. Figure 3 shows the facial feature points that we use in our research.  $\alpha_i$ , a vector expressing the coordinate of a feature point, is introduced and is described as:

$$\alpha_i = (x_i, y_i, z_i), \quad i = 1.1, 1.2, \dots, 5.3 \quad (1)$$

The origin of X-Y-Z coordinate system is assigned to be the tip of the nose. The information of the 23 facial feature points is used to calculate the 3D distances corresponding to the respective 3D faces. The coordinates of the facial feature points are transformed into a common coordinate system by subtracting the coordinates of the origin, denoted by the tip of nose (depicted as 4.1) . The transformed  $i^{th}$ , coordinate,  $\beta_i$  can be calculated by:

$$\beta_i = \alpha_i - origin, \quad i = 1.1, 1.2, \dots, 5.3 \quad (2)$$

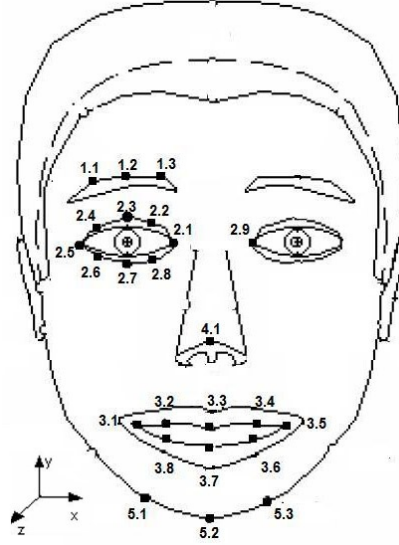


Figure 3: Facial feature points.

Each  $\beta_i$  value is normalized,  $\hat{\beta}_i$ , by dividing it by the width of the eye (distance between points 2.1 and 2.5) in order to compensate the distance effect between client faces. Thus we obtain the information  $\hat{\beta}_i$  for each facial expression of client.

$$\omega = \beta_{2.1} - \beta_{2.5} \quad (3)$$

$$\hat{\beta}_i = \beta_i / \omega \quad (4)$$

Moreover, in order to cancel out an individual variation and reflect the facial feature points movement more vividly, we calculate the 3D Distance Vectors depending on several facial feature points. Distance between two feature points is defined by:

$$(\hat{\beta}_i - \hat{\beta}_j) = \|\hat{\beta}_i - \hat{\beta}_j\| \quad (5)$$

Six different distances, which are used to form the distance vector for the facial expression recognition, are given from Eq.6 - Eq. 11. In Eq. 6, the



average of the three distances (between points 2.2–2.8, 2.3–2.7, and 2.4–2.6) is used in order to minimize the effect of a possible erroneous feature location that may contribute to the respective distance. This approach is adopted in Eq. 7, Eq. 8 and Eq. 11 respectively.

*Openness of eyes:*

$$\Delta_1 = \frac{1}{3}[(\hat{\beta}_{2.2} - \hat{\beta}_{2.8}) + (\hat{\beta}_{2.3} - \hat{\beta}_{2.7}) + (\hat{\beta}_{2.4} - \hat{\beta}_{2.6})] \quad (6)$$

*Height of eyebrows:*

$$\Delta_2 = \frac{1}{3}[(\hat{\beta}_{1.3} - \hat{\beta}_{2.2}) + (\hat{\beta}_{1.2} - \hat{\beta}_{2.3}) + (\hat{\beta}_{1.1} - \hat{\beta}_{2.4})] \quad (7)$$

*Openness of mouth:*

$$\Delta_3 = \frac{1}{3}[(\hat{\beta}_{3.2} - \hat{\beta}_{3.8}) + (\hat{\beta}_{3.3} - \hat{\beta}_{3.7}) + (\hat{\beta}_{3.4} - \hat{\beta}_{3.6})] \quad (8)$$

*Width of mouth:*

$$\Delta_4 = (\hat{\beta}_{3.5} - \hat{\beta}_{3.1}) \quad (9)$$

*Stretching of lip:*

$$\Delta_5 = (\hat{\beta}_{2.5} - \hat{\beta}_{3.1}) \quad (10)$$

*Openness of jaw:*

$$\Delta_6 = \frac{1}{3}[(\hat{\beta}_{5.1} - \hat{\beta}_{3.8}) + (\hat{\beta}_{5.2} - \hat{\beta}_{3.7}) + (\hat{\beta}_{5.3} - \hat{\beta}_{3.6})] \quad (11)$$

We transform the calculated distances into a set of Action Units (AUs) describing the facial expressions captured in the image sequence. We use a set of temporal rules and a fast 3D distance based procedure to encode 10 AUs occurring alone or in combination in an input face image sequence. Table 1 provides the list of the utilized rules.

Table 1: Rules for detecting AUs from a face image sequence. The value of  $\epsilon$  and  $\gamma$  have been decided based upon the threshold description provided by the relevant rule.

AU1 rule 1	Pulls the eyebrows' inner corners upward, causes the skin of the centre forehead to wrinkle horizontally. <i>IF <math>\Delta_2 &gt; \epsilon</math> AND <math>(\beta_{1.3} - \beta_{2.2}) &gt; \gamma</math> THEN AU1 is in action</i>
AU2 rule 2	Pulls the eyebrows' outer corners upward, causes the skin of the lateral forehead to wrinkle horizontally. <i>IF <math>\Delta_2 &gt; \epsilon</math> AND <math>(\beta_{1.1} - \beta_{2.4}) &gt; \gamma</math> THEN AU2 is in action</i>
AU4 rule 3	Pulls the eyebrows closer together, produces a bulge between the eyebrows, and lowers the eyebrows slightly. <i>IF <math>\Delta_2 &lt; \epsilon</math> AND <math>(\beta_{1.2} - \beta_{2.3}) &lt; \gamma</math> THEN AU4 is in action</i>
AU5 rule 4	Raises the upper eyelid and widens the eye opening. <i>IF <math>\Delta_1 &gt; \epsilon</math> AND <math>(\beta_{2.3} - \beta_{2.7}) &gt; \gamma</math> THEN AU5 is in action</i>
AU6 rule 5	Raises the cheeks and narrows the eye opening. <i>IF <math>\Delta_5 &lt; \epsilon</math> AND <math>\Delta_1 &lt; \gamma</math> THEN AU6 is in action</i>
AU7 rule 6	Raises the lower eyelid and narrows the eye opening. <i>IF <math>\Delta_1 &lt; \epsilon</math> AND <math>(\beta_{1.2} - \beta_{2.3}) &lt; \gamma</math> THEN AU7 is in action</i>
AU12 rule 7	Pulls the lip corners upward obliquely. <i>IF <math>\Delta_5 &lt; \epsilon</math> AND <math>\Delta_4 &gt; \gamma</math> THEN AU12 is in action</i>
AU17 rule 8	Pushes the chin boss and lower lip upward and stretches the skin on the chin boss. <i>IF <math>\Delta_6 &lt; \epsilon</math> AND <math>(\beta_{3.3} - \beta_{3.7}) &lt; \gamma</math> THEN AU18 is in action</i>
AU23 rule 9	Tightens the lips slightly making the lips appear narrower. <i>IF <math>\Delta_4 &gt; \epsilon</math> AND <math>(\beta_{3.3} - \beta_{3.7}) &lt; \gamma</math> THEN AU23 is in action</i>
AU27 rule 10	Parts the lips but does not stretch the mouth. <i>IF <math>\Delta_3 &gt; \epsilon</math> AND <math>\Delta_4 &lt; \gamma</math> THEN AU27 is in action</i>

### 3. Experiments

#### 3.1. AU Detection Results

For evaluating AU detection area underneath ( $A'$ ) the receiver-operator characteristic (ROC) curve is used. ROC curve is obtained by plotting true positive against the false alarm rate as the decision threshold varies. We have adopted a common statistic  $s = \sqrt{\frac{A'(1-A')}{\min\{n_p, n_n\}}}$  [14]. To maximise the amount of training and testing data, we have used leave-one-subject-out cross validation. We identify the thresholds of,  $\epsilon$  and  $\gamma$  for each AU rule from CK+ database [6]. The CK+ database consists of 593 FACS coded sequences from 123 subjects eliciting posed facial expressions. In our experiments we focused on the detection of AUs(1, 2, 4, 5, 6, 7, 12, 17, 23 and 27). The thresholds were selected that granted at true-positive rate of 80% and a false-positive rate of 10%. The results of AU detection for 3D distance vector based approach is given in Table 2. From the results it can be seen that proposed geometrical distance based features achieve very good detection accuracy. Even though the performance is slightly low for some AUs 4, 7 and 23, this is due to small variations in the geometrical orientations of the features.

#### 3.2. Affect recognition

The Inter-ACT corpus [15] has been developed to be a comprehensive repository of naturalistic and contextualised, task-dependent data for training and evaluation of an affect recognition system in an educational game scenario. The Inter-ACT corpus contains 156 six-second videos of the interaction between children and an iCat robot that play chess. It is an important advantage of the model that AUs are objective representations of human ex-

Table 2: Results showing the area underneath the ROC curve for 3D geometrical based features for AU detection.

Action Unit	Number of Occurrences	Detection Rate (%)
AU1	173	92.3 $\pm$ 2.2
AU2	116	95.8 $\pm$ 1.9
AU4	191	82.1 $\pm$ 3.1
AU5	102	97.6 $\pm$ 1.6
AU6	122	91.3 $\pm$ 2.7
AU7	119	80.3 $\pm$ 3.9
AU12	111	96.3 $\pm$ 2.1
AU17	196	90.1 $\pm$ 1.9
AU23	59	82.3 $\pm$ 4.3
AU27	81	97.3 $\pm$ 1.1

pressions and are independent from any assigned interpretation, thus allowing further high level decisions and processing.

In many application domains the knowledge about the view direction of the eyes is more important than the orientation of the head, respectively the face. But the measurements relying on the eyes and the head are usually related to each other. For simplicity and reducing computational complexity we defined gaze-tracking as head-direction recognition. By recognising where a users head is directed we can infer the direction of their gaze. This may be sufficient to identify whether they are looking at the game companion in our 3D environment or distinguish between gaze directed at the chess board or elsewhere. This means that the system knows very accurately where the user is looking and when they are looking there. In the proposed framework,

Table 3: Affect descriptions in terms of facial action units and head direction

Affect State	AU criteria	Head Direction
Engaged	{AU6,AU12}	Looking at iCat
Avoidance	{AU1,AU2,AU5,AU27} or {AU1,AU2,AU4}	Looking at iCat
Aggressive	{AU4,AU5,AU7}	Looking at iCat
Calm	{AU17} or {AU23} or No AUs in action	Looking at iCat
Pleasant	{AU12}	Looking at board
Confused	{AU1,AU2,AU4} or {AU5}	Looking at board
Unpleasant	{AU4} or {AU17}	Looking at board
Thinking	{AU23} or No AUs in action	Looking at board
Relax	{AU12}	Looking elsewhere
Not Engaged	{AU5} or {AU27}	Looking elsewhere
Tense	{AU4}	Looking elsewhere
Tired/Bored	{AU17} or {AU23} or No AUs in action	Looking elsewhere

the FACS model and direction of the head are used in real-time to detect complex behavioural states as listed in Table 3.

Furthermore, we divided each facial action into four temporal segments: the neutral(no action), onset(beginning), apex(peak), and offset(ending) as depicted in Figure 4. We define each temporal rule for AU detection in a unique way according to relevant rule and using statistical information of the last 8 consecutive frames of the image sequence. To minimize the the effects of noise and inaccuracies in facial feature tracking and to enable detection of the temporal dynamics of displayed AUs, we consider coefficient of variance,  $C$  (e.g.  $C > 0.05$  indicates variance in the 3D distance vector is greater than 5%), for each 3D distance vector. We identify the thresholds of  $\rho$ , for each AU rule from Inter-ACT corpus. Incited by the research findings that

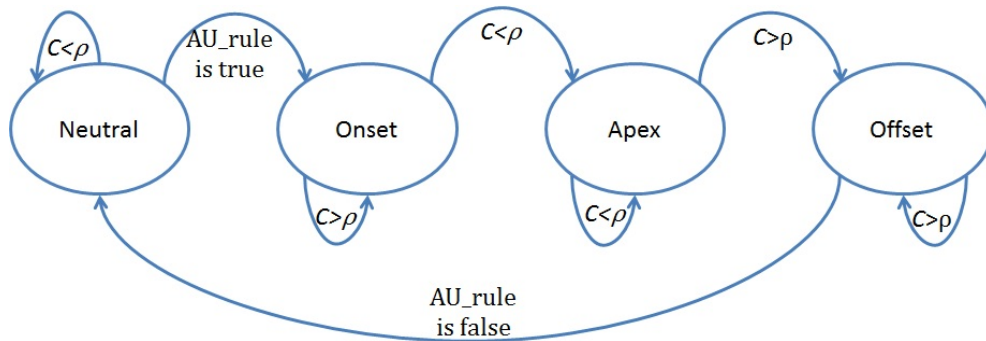


Figure 4: Temporal rules for AU detection

suggested that temporal changes in neuromuscular facial activity may last from 1/4 of a second to several minutes [11], the temporal domain has been determined empirically based on a video frame rate of the input sequence (i.e., 8 frames approximately have a duration of 1/4 of a second for 30 fps).

The confusion matrix of the average case for the user looking at the game companion is 95.9% as shown in Table 4. Note that most of the expressions are detected with high accuracy and the confusion is larger with the Aggressive and Calm behaviours. One reason why Aggressive is detected with only 92.8% is that in general these behaviour is eyebrow dependent and confusion with Calm is much larger than with the other behaviour as one can easily observe from naturalistic face images. Results for user looking at the chess board is presented as a confusion matrix in Table 5. It can be seen that the highest misclassification occurs between the expressions of Unpleasant and Thinking. The decrease in these recognition rates is attributed to their similarity especially in low intensities. The main difference between the Unpleasant and Thinking behaviours lies mostly on the configuration of the eyebrows, which cannot be effectively captured using depth (at least with

Table 4: Confusion matrix of affect recognition for “Looking at the iCat”

	Engaged	Avoidance	Aggressive	Calm
Engaged	98.6	0.0	0.0	1.4
Avoidance	0.0	100.0	0.0	0.0
Aggressive	0.0	0.0	92.8	7.2
Calm	1.2	0.0	6.7	92.1

Table 5: Confusion matrix of affect recognition for “Looking at the chess board”

	Pleasant	Confused	Unpleasant	Thinking
Pleasant	96.7	0.0	0.5	2.8
Confused	0.0	99.3	0.0	0.7
Unpleasant	0.0	2.1	88.5	9.4
Thinking	1.7	0.0	8.1	90.2

our point correspondence technique) especially in low intensities, where the difference is so subtle even for a human eye. Confusing matrix for user looking elsewhere is shown in Table 6. Since both Tense and Tired/Bored tighten the lips causes misidentification of Tired/Bored when the mouth corners are tracked. Note that Tense and Tired/Bored are also often confused by humans. Hence, the distinction between these two behaviours may be more amenable to appearance features than to geometrical features.

#### 4. Conclusions and Future Work

In this paper, we have presented an initial computational model for the recognition of engagement with a robotic game companion in an educational scenario. The proposed approach is based on the automatic analysis of af-

Table 6: Confusion matrix of affect recognition for “Looking elsewhere”

	Relax	Not Engaged	Tense	Tired/Bored
Relax	99.1	0.0	0.6	0.3
Not Engaged	0.0	99.2	0.0	0.8
Tense	0.0	5.4	82.1	12.5
Tired/Bored	1.2	2.9	9.2	86.7

fective states. Differently from many systems proposed in the literature, our approach is based on vision-based on real-time extraction of facial features from videos capturing users behaviour from a non-posed facial behaviour in naturalistic environment. This allows for the dynamics of facial behaviour to be analysed in a more comprehensive manner. The evaluation conducted in the same interaction scenario of the final application suggests that patterns of facial behaviour with the head direction information can be used to accurately predict the engagement of the children with the robot. Experimental results highlighted the key role played by the temporal dynamics of neuromuscular actions in automatic engagement recognition. The high recognition accuracy achieved with the computational model makes our approach suitable for integration into an affect recognition system for a game companion in a naturalistic scenario.

Future work to further validate the proposed approach will include a more comprehensive evaluation with a larger number of video samples to train the automatic recognition models, as well as the design of a framework for fusion with other modalities (e.g., contextual information) of interest for this scenario.



## 5. Acknowledgments

This work was supported by the EU FP7 ICT-215554 project LIREC (Living with Robots and intERactive Companions).

## References

- [1] M. Pantic, N. Sebe, J. F. Cohn, T. Huang, Affective multimodal human-computer interaction, in: Proceedings of the 13th annual ACM international conference on Multimedia, MULTIMEDIA '05, ACM, New York, NY, USA, 2005, pp. 669–676.
- [2] F. De la Torre, T. Simon, Z. Ambadar, J. F. Cohn, Fast-facs: a computer-assisted system to increase speed and reliability of manual facs coding, in: Proceedings of the 4th international conference on Affective computing and intelligent interaction - Volume Part I, ACII'11, Springer-Verlag, Berlin, Heidelberg, 2011, pp. 57–66.
- [3] Z. Zeng, M. Pantic, G. Roisman, T. Huang, A survey of affect recognition methods: Audio, visual, and spontaneous expressions, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31 (2009) 39–58.
- [4] A. Vinciarelli, M. Pantic, H. Bourlard, Social signal processing: Survey of an emerging domain, *Image and Vision Computing Journal* 27 (2009) 1743–1759.
- [5] K. R. Scherer, *Psychological models of emotion*, Oxford University Press, Oxford/New York, j. borod (ed.) edition, pp. 1337–166.

- [6] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression, in: Workshop on Human Communicative Behaviour Analysis in conjunction with CVPR.
- [7] R. S. Aylett, G. Castellano, B. Raducanu, A. Paiva, M. Hanheide, Long-term socially perceptive and interactive robot companions: challenges and future perspectives, in: Proceedings of the 13th international conference on multimodal interfaces, ICMI '11, ACM, New York, NY, USA, 2011, pp. 323–326.
- [8] M. Yeasin, B. Bulot, R. Sharma, Recognition of facial expressions and measurement of levels of interest from video, *IEEE Transactions on Multimedia* 8 (2006) 500–508.
- [9] A. Yazdani, J.-S. Lee, J.-M. Vesin, T. Ebrahimi, Affect recognition based on physiological changes during the watching of music videos, *ACM Trans. Interact. Intell. Syst.* 2 (2012) 7:1–7:26.
- [10] A. J. N. van Breemen, X. Yan, B. Meerbeek, icat: an animated user-interface robot with personality, in: 4th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2005), July 25-29, 2005, Utrecht, The Netherlands, pp. 143–144.
- [11] P. Ekman, W. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement., Consulting Psychologists Press, Palo Alto, 1978.

- [12] H. Soyel, H. Demirel, Facial expression recognition using 3d facial feature distances, in: M. Kamel, A. Campilho (Eds.), *Image Analysis and Recognition*, volume 4633 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, 2007, pp. 831–838.
- [13] H. Soyel, H. Demirel, 3D facial expression recognition with geometrically localized facial features, in: *Proceedings of the 23rd International Symposium on Computer and Information Sciences, ISCIS'08*, pp. 1–4.
- [14] C. Cortes, M. Mohri, Confidence Intervals for the Area under the ROC Curve, in: *Advances in Neural Information Processing Systems (NIPS 2004)*, volume 17, MIT Press, 2005.
- [15] G. Castellano, I. Leite, A. Pereira, C. Martinho, A. Paiva, P. W. McOwan, Inter-act: an affective and contextually rich multimodal video corpus for studying interaction with robots, in: *Proceedings of the international conference on Multimedia, MM '10*, ACM, New York, NY, USA, 2010, pp. 1031–1034.