Micro-Expression Recognition for Detecting Human Emotional Changes

Kazuhiko Sumi(✉) and Tomomi Ueda

Aoyama Gakuin University, Sagamihara, Kanagawa 252-5258, Japan
sumi@it.aoyama.ac.jp

Abstract. We propose a method estimating human emotional state in communication from four micro-expressions; mouth motion, head pose, eye sight direction, and blinking interval. Those micro-expressions are picked up by a questionnaire survey of human observers watching on video recorded human conversation. Then we implemented a recognition system for those micro-expressions. We detect facial parts from a RGB-Depth camera, measure those four expressions. Then we apply decision-tree style classifier to detect some emotional state and state changes. In our experiment, we gathered 30 videos of human communicating with his/her friend. Then we trained and validated our algorithm with two-fold cross-validation. We compared the classifier output with human examiners' observation and confirmed over 70 % precision.

1 Background and Objectives

In recent years, face recognition application to communications media and human interface, as well as research for human face recognition has been active in computer science. Studies of facial expression recognition by computer system, especially technique for still face image expression recognition, have been popular in these 15 years. More recently, studies have proceeded toward video face images.

Quantitative description of facial expression was first studied by Ekman [1]. He developed face behavioral description method, which is referred to as Facial Action Coding System (FACS). In FACS, the face area is divided into three areas; top: around eyebrow, central: around eye, bottom: around mouth. In those three areas, he defined standard unit movement of facial parts, in other word the movement of the muscles of the face, which is referred to as Action Unit (AU). AUs were classified into 44 types. Human six major facial expression, "happiness", "fear", "dislike", " surprise", "sadness", and "anger" are described by the combination of several AUs.

However, the above-mentioned six major expressions are somewhat very distinctive, intentionally posed expression. In our daily communication, natural facial expression is more subtle and delicate. It is so called micro-expression. Micro-expression, also explained by Ekman, is the rapid change of facial expression and appears only in a short-period. However, detailed description of micro-expression

M. Kurosu (Ed.): HCI 2016, Part III, LNCS 9733, pp. 60–70, 2016.
DOI: 10.1007/978-3-319-39513-5_6

or relationship between emotion and micro-expression is not yet established. Current studies are focusing on searching clue for estimation of emotional state, not limited to the face, speech and body motion, and voice.

In this study, we look for facial motion and head motion that becomes a clue for emotion estimation appeared in human-to-human conversation and communication. We described the useful micro-expression and relationship between emotion and micro-expression from the analysis of questionnaire survey of observers watching on human conversation videos. We implemented the observer's analysis into computer and compared its estimation with human observer's one. Although, it is a subjective judgment and there is no evidence that the analysis is exact to the mental state of test subjects, there is correlation between human observer's estimation and computer outputs. We expect this system can be applied to machine-to-human communication that have the power of empathy and warm atmosphere.

2 Related Work

Most of the existing studies on human emotion recognition are based on automatic facial expression recognition. Ekman and Friesen developed the Facial Action Coding System (FACS)[1]. 44 facial action units (AU) are defined to describe facial expression. Basic emotions, i.e., happiness, surprise, anger, sadness, fear, and disgust are corresponding to prototypic facial expressions.

Based on this idea, many studies of automatic recognition of prototypic facial expressions were carried out. For example, Black [3] detected prototypic expressions by combination of facial parts motion and deformation. Facial parts are detected by facial part image templates. Mase [2] detected prototypic expressions from optical flow on the face image. Essa [4] detected facial control points and detected AUs from the motion of control points. Donato compared the performance of several features and classification approach i.e., optical flow, PCA, LFA, FLD, ICA, and their local patch versions. They also compared automatic method with human estimation. He concluded that ICA and Gabor Jets based approach are the best performance. However, images are frontal face image, cropped, normalized, and marked manually. Those studies were principle but cannot be applied to real applications. Tian et.al., developed multi-state feature-based AU recognition [6]. Their method could recognize non-frontal faces if all the AU muscles were seen. However, there is a manual facial feature point refinement process and it is not fully automatic.

So far, face analysis methods are classified into three types according to the facial features. First is geometric feature (shape of facial parts) approach. For example, Chang et.al., used 58 facial landmarks [8]. Second is facial feature point based approach. For example, Pantic et.al., used facial characteristic points around facial parts [9]. Third is facial texture features approach. For example Bartlett et.al., uses Gabor wavelets to describe facial shape changes [7]. More recently, those features are integrated and the precision of recognition is improved [13,14].

There are few approaches that integrate information from facial expressions and head motion. Body motion, such as head pose or hand gesture is more visible rather than a small change of facial expression, and appears to be corresponding to a certain emotional state. For example, Asteriadis et.al, integrated eye gaze state and head pose to describe e-learners' behavior [10]. Gunes combined facial expressions and body motion to estimate human affect recognition [12].

In these 5 years, human-to-machine interface, usage of RGB-Depth (RGBD) camera has become popular. Compared with standard RGB camera, advantages of RGBD camera for face recognition is robustness and performance [15]. However, most of the studies remain in basic study, and emotion recognition in a natural conversation environment is still a big challenge. In this study, we aim at finding useful emotion categories and corresponding facial/body expressions appearing in hume-to-human communication in real world situation.

3 Proposed Method

In the conventional technique, facial expressions that have been recognized are obvious expressions, while a human in daily life read more delicate emotions. To realize such delicate emotion recognition, we utilize not only obvious facial expression but also micro-expressions and other body motion. Micro-expression is that of the moment appear relatively natural facial expressions and facial behavior in expression. Detecting micro-expression is considered to be important information on changes in the human delicate emotions. Therefore, rather than the change of motion in three areas used in the conventional AU technique (eyebrows, eyes, and mouth), we look for new micro-expression AUs for micro-expressions.

To simplify the problem, we focus on estimating emotion during communication or conversation, and try to find several emotion classes that can be stably observed by both human and computer.

To find such emotion class, we conducted a preliminary experiment with 10 test human observers. First, we recoded video of a person in communication and showed it to 10 observers. Each observer was asked to describe what kind of emotions he/she estimated about the person in the video. By analyzing all the observers' description and fining common descriptions, we come to a conclusion that the following five emotions are appearing in conversations; "friendly[1]", "boring", "a little depression", "shocked", and "a little surprised".

Then we looked for corresponding micro expression to those five emotions. We showed the video to the observers again and asked to describe which of the five emotions he/she discovered and the clue why he/she discovered the emotion. By analyzing all the observers' description again, we correlated the following micro-expression to the five emotions; mouth motion, face direction, eye sight direction, and blinking interval. Details are described in Sect. 3.1.

[1] We define friendly is a mental attitude attracted by the partner's talk or the partner him/herself.

We implemented face parts recognition and above emotion estimation with RGBD camera images. Figure 3 shows the schematic diagram of our proposed method. (Figs. 1, 2, 4, 5, 6, 7, 8, 9 and 10)

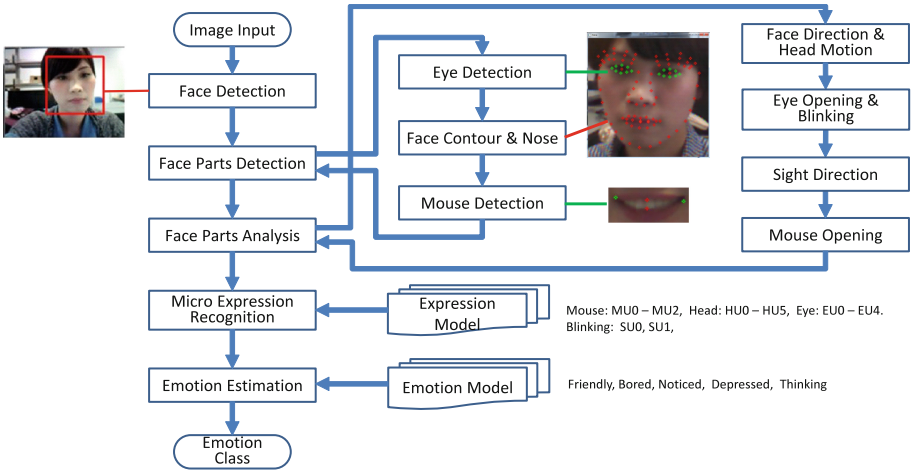

Fig. 1. Schematic Diagram of Emotional Estimation

The input is a pair of a 3D range and a RGB texture image of a human, taken by RGBD camera. First, the face region is detected by a combination of depth peak and facial pattern. In the figure, red rectangle in the left picture denotes the detected face region. Then eye and face contour are detected using the face texture model and the depth edge. In the figure, green dots on the right picture denote eye contours and red dots denote face contour and other facial parts contour. Once, facial parts and their locations are detected, we perform measurement of facial components; head direction, eye opening and blinking, line of sight direction, and mouth opening. Those measurements are matched with micro-expression model, which is built from the observers questionnaires. Finally, we estimate emotions and their changes from the emotion model.

Mouth Motion. According to emotional condition, mouth open width and stretched length are changing variously. For example, laughing is a obvious action. Laugh opens the mouth widely and the teeth are disclosed. On the other hand, smile, which is more delicate expression than laugh, raises the corner of the mouth just a little. In conversation, the mouth is changing its shape variously to speak. Thus it is not perfect estimating emotions only from mouth motion. Never the less mouth motion is very important information for estimating delicate emotions.

We focus on mouth open width, which is the distance between lower edge of the upper lip and upper edge of the lower lip m_y, and mouth stretch length, which

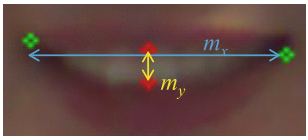

Fig. 2. Detecting mouth motion; opening width and stretching length

is the distance between the left and the right corner of the mouth m_x. If m_x, m_y exceeds a pre-determined threshold, we detect the following three motions; smiling (MU0), mouth slightly opening (MU1), and mouth closing (MU2) as in Eq. 1.

$$\mathrm{MU} = \begin{cases} 0 \text{ (smile)} & \text{if } m_x \geq T_{MUx1} \text{ and } m_y \leq T_{MUy1} \\ 1 \text{ (slightly open)} & \text{if } m_x \geq T_{MUx2} \text{ and } m_y \geq T_{MUy2} \\ 2 \text{ (close)} & \text{if } m_x \geq T_{MUx3} \text{ and } m_y \leq T_{MUy3} \end{cases} \tag{1}$$

where $T_{MUx1} = 0.22$, $T_{MUy1} = 0.00$, $T_{MUx2} = 0.19$, $T_{MUy2} = 0.05$, $T_{MUx3} = 0.17$, and $T_{MUy3} = 0.01$ of horizontal face size in our implementation.

Head Pose. Psychologists pointed out that lowers his head when sad and body tremble when he is scary. Empirically, we know the strong evidence that emotion affects head post. For example, head rotates naturally its direction toward an interested object or person. Head pose go up when feeling contemptuous of a person. Head pose go down when feeling shame, sadness, embarrassment, and bored. This is a non-verbal communication of "attitude", when we are in conversation.

We compute an average facial surface normal from a range image of the face region of the subject. Then compute three face directional angles; roll θ_x (rotation around X axis), pitch θ_y (rotation around Y axis), and yow θ_z (rotation around Z axis) respectively. Figure 3 shows the axis of the head. If those angle exceeds a pre-determined thresholds, we detect 6 face directional motion; directing front (HU0), directing left (HU1), directing right (HU2), directing up (HU3), directing down (HU4) and nodding (HU5) as in Eq. 2.

$$\mathrm{HU} = \begin{cases} 0 \text{ (front)} & \text{if } \theta_x \leq T_{HUx1} \text{ and } \theta_y \leq T_{HUy1} \text{ and } \theta_z \leq T_{HUz1} \\ 1 \text{ (left)} & \text{if } \theta_y \geq T_{HUy2} \\ 2 \text{ (right)} & \text{if } \theta_y \leq -T_{HUy2} \\ 3 \text{ (up)} & \text{if } \theta_x \geq T_{HUx2} \\ 4 \text{ (down)} & \text{if } \theta_x \leq T_{HUx3} \\ 5 \text{ (nodding)} & \text{if } T_{HUx4} \leq \theta_x \leq T_{HUx5} \end{cases} \tag{2}$$

where $T_{HUx1} = 9$, $T_{HUy1} = 8$, $T_{HUz1} = -3$, $T_{HUy2} = 24$, $T_{HUx2} = 15.6$, $T_{HUx3} = -8$, $T_{HUx4} = -7$, and $T_{HUx5} = -3$ degree in our implementation.

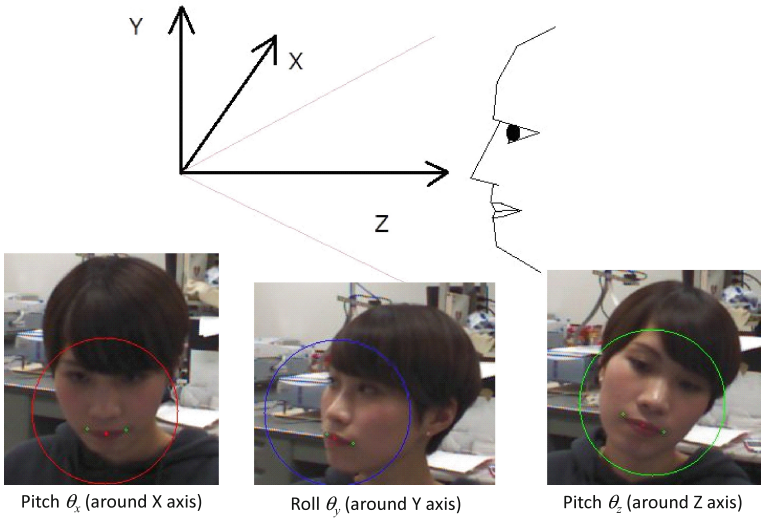

Fig. 3. Recognition of face/head pose (Color figure online)

Direction of Line of Sight. Among the expression, in particular eye produces significant and direct impression. We naturally feel that information from thoughtful eyes and their motion is equivalent to spoken words. Sometimes we are able to distinguish posed smile from a laughing face. This is because we are reading the movement of eyes. For example, if the line of sight is looking up, it implies remembering with the past experience or the landscape as seen previously. If eyes and facing up is moving left and right restlessly, it implies upset feelings. So, eye movements often represent the feelings unconsciously.

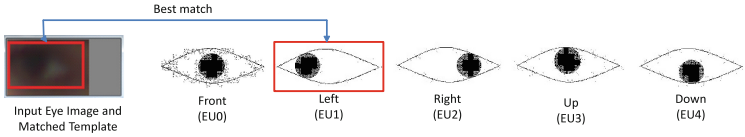

Fig. 4. Recognition of line of sight direction.

To detect line of sight or eye direction, we first detect each eye region (left and right). Then for each eye, we compare the region with 5 typical pre-determined template images expressing looking front (EU0), looking left (EU1), looking right (EU2), looking up (EU3) and looking down (EU4) as in Fig. 3 and Eq. 3. In Fig. 3, the left image is the detected eye region, red rectangle denotes highest match among 5 templates. In this case, template EU1 is the best match.

$$\mathrm{EU} = \arg\max_{0 \le k \le 4} \max\{q(I(x,y), G(k))\} \tag{3}$$

where, k is the template number ($0 \leq k \leq 4$), $G(k)$ is the k-th template image (size $w \times h$), ($I(x, y)$ is a ($w \times h$) sub-region of each eye region starting from upper left corner at (x, y), and $q(I, G)$ is a correlation function. Our implementation uses normalized cross-correlation function as q.

Blinking. Blinks, usually 25 to 35 times per minutes, become more than 35 times, when an impact is applied to the eye or sudden emotional influence, such as upset and surprises. Blinking or frequency has a good correlation with mental state whether the person is nervous or relaxed.

To measure eye opening and blinking, we use skin color based approach. We compute the ratio of dark color (iris and pupil region) pixels and skin color pixels of the eye region of the test subject, then estimate eye opening and closing comparing with a pre-determined threshold. Figure 3 shows the scheme. In the figure, left upper image is the detected eye region, left middle and left bottom images are eye opened and closed image respectively. White pixel denotes that color is similar to skin. If the number of skin color pixels exceeds the threshold, we count one eye closing (pointed by red arrows in the figure). Then we count the number of eye closing in a few seconds and compute blinks per minutes n_b. If n_b is less than a threshold, we consider it is stable (SU0), if it is larger we consider it is nervous (SU1) as in Eq. 4.

$$\text{SU} = \begin{cases} 0 \text{ (stable)} & \text{if } T_{SU1} \leq n_b \leq T_{SU2} \\ 1 \text{ (nervous)} & \text{if } n_b > T_{SU2} \end{cases} \tag{4}$$

where, threshold $T_{SU1} = 25$ and $T_{SU2} = 35$ in our implementation.

3.1 Emotion Estimation

To build a mental state corpus, we recorded 30 cut of video of the test subject conversations with his/her friend. Then we showed the video to the evaluator subjects, and asked why they felt the five emotions. From their answers, we could find corresponding micro-expressions related with the five emotions as in Table 1.

Table 1. Correspondence between micro-expressions and emotions

	Eye-sight	Blinking	Face direction	Mouth motion
Feeling familiar	EU0	SU0	HU5	MU0
Bored	EU1 $\vee$ EU2	SU1	HU1 $\vee$ HU2	MU2
Noticing	EU0	SU1	HU3	MU1
Depressed	EU4	SU0	HU4	MU2
Thinking	EU3	SU0	HU3	MU2

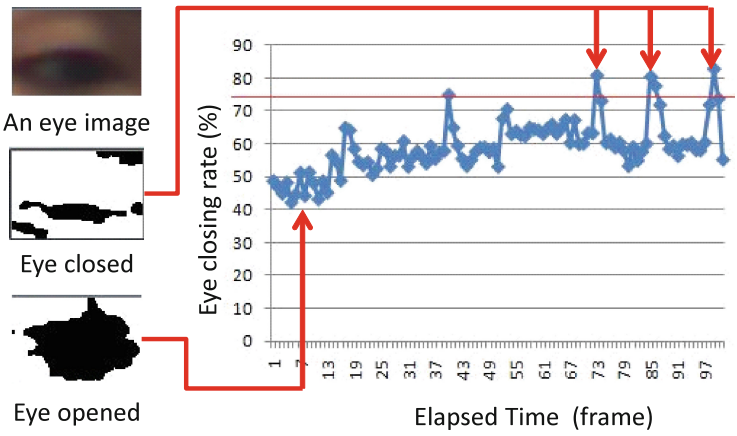

Fig. 5. Recognition of eye opening and blinking

According to the co-occurrence of the micro-expressions for each emotions in Table 1, we can build a decision tree, in which each node of the tree corresponds to a row of Table 1. If there is a match of AUs combination at a node, corresponding emotion is detected. (Tables 2 and 3)

Table 2. Emotional Transition and Corresponding Facial Action Units

emotion	Combination of Micro-expression	
Feeling Familiar	EU0 ∧ HU5 ∧ MU2	
Boredom	EU1 ∧ HU1	EU2 ∧ HU2
Noticed	EU0 ∧ HU5 ∧ MU1	SU1 ∧ MU1
Depressed	EU3 ∧ HU4	
Thinking	EU0 ∧ HU3	

4 Experiments

Using 30 video cuts generated from our video corpus, we performed two-fold cross-validation. We trained our system with half of the corpus. Then we examined the rest of the corpus for evaluation. We asked 10 experimenters, different persons from observers in the preliminary experiment in Sect. 3, to check a conversation video, in which each of scene cuts contains a single emotional expression. Then we evaluated computer's output with the human experimenters' results. The results are shown in Table 4.

As the second experiment, we aimed to detect multiple emotional expressions. In this case, experimenter (same as previous experiment) are asked to

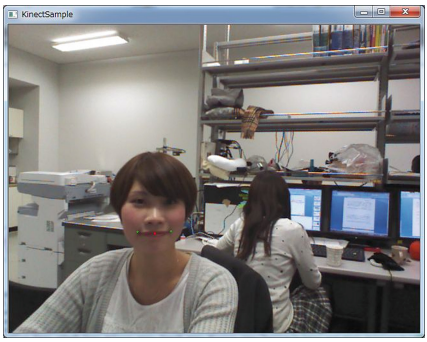

Fig. 6. An example scene for single emotion evaluation

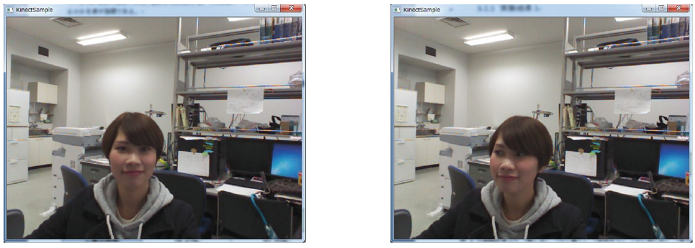

Fig. 7. Examples of friendly emotion estimation. Successfully estimated (left) and failure (right)

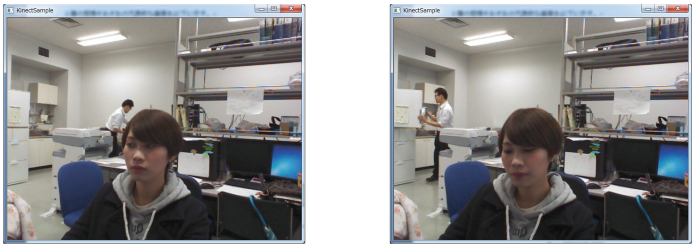

Fig. 8. Examples of bored emotion estimation. Successfully estimated (left) and failure (right)

Fig. 9. Examples of thinking estimation. Successfully estimated (left) and failure (right)


Fig. 10. Examples of noticed emotion estimation. Successfully estimated (left) and failure (right)

Table 3. Single emotion in a cut

Single Emotion	Recall Rate	Precision
Single Emotion	83 %	80 %

check if there is a transition of emotion. This means that in the first half of the video contains emotion A, while the second half of the video contains another emotion B. Of course it is more difficult task, because the algorithm as well as the experimenter have to estimate two emotion correctly. The results are shown in Table 4.

Table 4. Multiple emotions in a cut

Emotional Change	Recall Rate	Precision
Emotion A to B	74 %	70 %

Table 4 shows that a single emotion can be estimated more the 80 % from our method. This means that the facial parts recognition and micro-expression recognition proposed in Sect. 3.1 is working well and corresponding emotional state estimation is working too. However, Table 4 shows emotional changes is about 10 % less accurate than single emotion. This implies that our method is somewhat different from human estimation. We found that human evaluators have a tendency to feel continuous even after the first emotional cue distinguished. We should develop some hysteresis function in the future.

5 Conclusion

We proposed a method to micro-expression recognition for detection of human emotional changes. We focused mouth motion, face direction, eye-sight direction, and blinking as micro-expressions. Our experiments showed a good corresponding between our method and human evaluators. However, our method has some difference when human shows multiple emotions or changes their emotions.

References

1. Ekman, P., Friesen, M.V.: The Facial Action Coding System: A Technique for The Measurement of Facial Movement. Consulting Psychologist, Palo Alto (1978)
2. Mase, K.: Recognition of facial expression from optical flow. IEICE Trans. **E74**(10), 3474–3483 (1991)
3. Black, M., J., Yacoob, Y.: Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion. In: International Conference on Computer Vision, pp. 374–381 (1995)
4. Essa, I., Pentland, A.: Coding, analysis, interpretation and recognition of facial expressions. IEEE Trans. PAMI **19**(7), 757–763 (1997)
5. Donato, G., Bartlett, M.S., Hager, J.C., Ekman, P., Sejnowski, T.J.: Classifying facial actions. IEEE Trans. PAMI **21**(10), 974–989 (1999)
6. Tian, Y., Kanade, T., Cohn, J.F.: Recognizing action units for facial expression analysis. IEEE Trans. PAMI **23**(2), 1–19 (2001)
7. Bartlett, M.S., Littlewort, G., Frank, M.G., Lainscsek, C., Fasel, I., Movellan, J.: Recognizing facial expression: machine learning and application to spontaneous behavior. In: IEEE Conference on CVPR, pp. 568–573 (2005)
8. Chang, Y., Hu, C., Feris, R., Turk, M.: Manifold based analysis of facial expression. J. Image Vis. Comput. **24**(6), 605–614 (2006)
9. Pantic, M., Patras, I.: Dynamics of facial expression: recognition of facial actionsnand their temporal segments from face profile image sequence. IEEE Trans. SMCB **36**(2), 433–449 (2006)
10. Asteriadis, S., Tzouveli, P., Karpouzis, K., Kollias, S.: Estimation of behavioral user state based on eye gaze and head pose – application in an e-learning environment. Multimedia Tools Appl. **41**(3), 469–493 (2008)
11. Gunes, H., Piccardi, M.: Automatic temporal segment detection and affect recognition from face and body display. IEEE Trans. SMC. Part B, Cybern. **39**(1), 64–84 (2009)
12. Gunes, H., Pantic, M.: Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners. In: Safonova, A. (ed.) IVA 2010. LNCS, vol. 6356, pp. 371–377. Springer, Heidelberg (2010)
13. Bartlett, M.S., Whitehill, J.: Automated facial expression measurement: recent applications to basic research in human behavior, learning, and education. In: Calder, A., et al. (eds.) Handbook of Face Perception. Oxford University Press, New York (2010)
14. Tian, Y., Kanade, T., Cohn, J.F.: Facial expression recognition. In: Li, S.Z., Jain, A.K. (eds.) Handbook of Face Recognition, pp. 487–520. Springer-Verlag, Berlin (2011). Chap. 11
15. Lemaire, P., Ardabilian, M., Chen, L., Daoudi, M.: Fully automatic 3D facial expression recognition using differential mean curvature maps and histograms of oriented gradients. In: International Conference on Automatic Face and Gesture Recognition, pp. 1–7 (2013)