ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

FACIAL EXPRESSSION RECOGNITION AND THEIR TEMPORAL SEGMENTS FROM FACE PROFILE IMAGE SEQUENCES

Mr. Easari Parusha Ramu, Assistant Professor, Department Of ECE SICET Hyderabad

Navya Sree Bandi, Pavan Bonala, Snehith Reddy Challa, Umesh Chandra Dharavath

UG Student, Department Of ECE, SICET, Hyderabad

ABSTRACT

Temporal segmentation of facial movements in spontaneous facial movements recorded in real-world settings is an important open and relatively unknown problem in facial image analysis. Several issues contribute to the challenge of this task. These include non-frontal postures, moderate to large out-of-plane head movements, large variations in the time scale of facial movements, and the exponential nature of possible combinations of facial movements. To address these challenges, we propose a two-step approach for temporal segmentation of face actions. The first step uses spectrogram techniques to cluster shape and appearance features that are invariant to several geometric transformations. The second step groups the clusters into temporally constant facial expressions. We evaluated our method on facial movements recorded during face-to-face interactions. Video data were originally collected to answer fundamental questions in psychology that were not related to algorithm development. This method achieved moderate convergent validity using manual FACS (Facial Action Coding System) annotation. Furthermore, using this method to pre-process videos for manual FACS annotation can significantly increase productivity and eliminate the need for ground-truth data for face image analysis. In addition, we were able to detect abnormal facial movements.

INTRODUCTION

Temporal segmentation of facial actions from videos is an important open problem in automatic face image analysis. With few exceptions, previous literature treated video frames as if they were independent and ignored their temporal organization. Facial movements have an onset, one or more peaks, and an offset, and the temporal organization of these events is important for understanding and understanding facial expressions [2, 5, 6]. For automatic face image analysis, temporal segmentation is important for decomposing facial movements into action units (AUs) and higher orders.

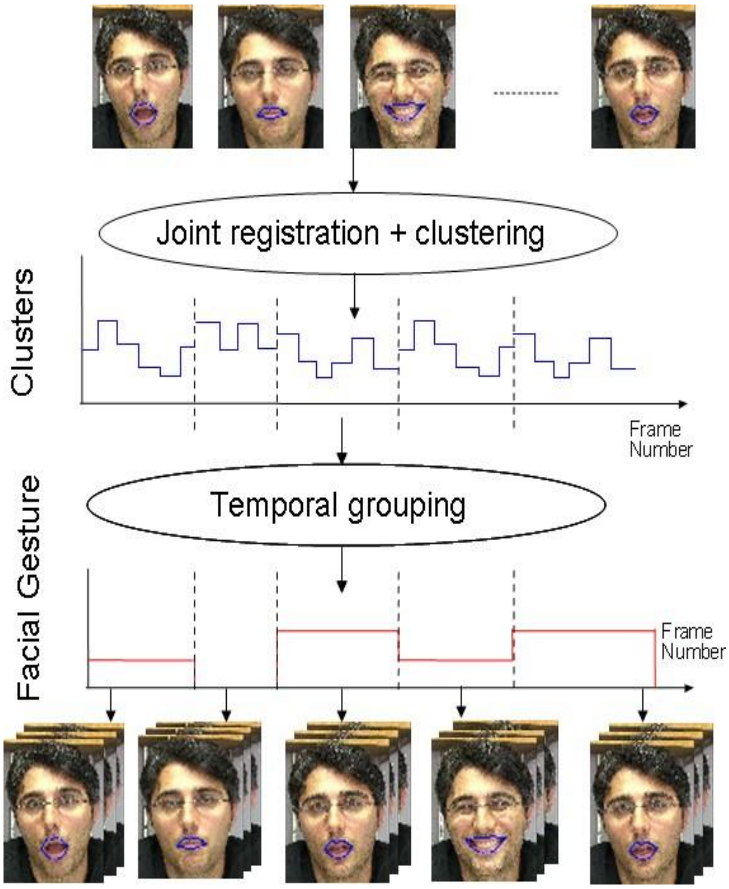


Figure 1. Temporal segmentation of facial gestures.

combinations or expressions [5], to improving recognition

For applications such as facial expression recognition function, abnormal facial expression detection is used.

Several factors make it difficult to recover the temporal structure of facial expressions from videos, especially in real-world environments characterized by non-frontal poses, moderate out-of-plane head movements, subtle facial movements, large movements, etc. A tough challenge when video is acquired in a typical environment. Variation in the time scale of facial actions (both within and between event classes) and the exponential number of possible facial action combinations. To address these challenges, we propose a two-step approach for temporal segmentation of face actions. The first step uses spectrogram techniques to cluster shape and appearance features. The resulting clusters are invariant to several geometric transformations. The second step groups the clusters into temporally constant facial expressions (Figure 1).

This article is organized as follows. Section 2 reviews previous works on facial expression detection and temporal segmentation. Section 3 reviews the state of the art in clustering algorithms. Section 4 proposes a method to discover temporal clusters. Section 5 presents experimental results and two new applications of our method. Recognize unusual or abnormal facial behavior. Another improves the efficiency of manual FACS annotation by video preprocessing using automatic segmentation.

SYSTEM REVIEW

During the last decade, considerable efforts have been devoted to automatic face image analysis. Major topics include facial feature tracking, facial expression analysis, and facial recognition [36, 26, 22]. Facial expressions refer to both emotional states (such as happy or sad) and anatomically based facial movements [14]. Comprehensive reviews of automated facials can be found in [30, 36, 22, 33]. Here we briefly review the literature related to the current study.

The pioneering work of Black and Yacoub [3] was recognized by fitting a local parametric motion model in the face region and feeding the resulting parameters to the nearest neighbor classifier for facial expression detection. De la Torre et al. [10] using density and appearance models,

belongs to class Cj, c is the number of classes and n is the number of samples. The D × n columns contain the original data points and the M columns represent the cluster centers. d is the dimension of data. Equivalence of the K-means error function and Eq. 2 is valid only if G strictly satisfies the constraints.

The K-means algorithm performs coordinate descent on E1 (M, G). Given the true value of the mean M, the first step is to minimize the expression for each data point dj by finding gj such that one column is 1 and the other column is 0.

2. The second stage is optimized over M = DG(GT G)-1. This is equivalent to calculating the average of each cluster. Although we can prove that the alternation of these two steps always converges, the K-means algorithm does not always find the optimal configuration in all possible assignments. Typically, the algorithm is run several times from different initial conditions and the best solution is selected. Despite its limitations, this algorithm is easy and effective to implement, so it is often used.

One of the advantages of relating a clustering problem to an error function is that the bounds can be easily derived. For example, after removing M, the formula becomes: 2 can be rewritten as:

E2(G) = ||D − DG(GT G)−1GT ||F = tr(DT D)

T 1 T T min(d,n) i=c+1

Here, λi are the eigenvalues of DT D. Minimize expression. 3 is equivalent to maximizing tr((GT G) -1GT DT DG). Ignoring the specific structure of G and considering the continuous domain, the value of G that minimizes Eq. 3 is given by the eigenvectors of the covariance matrix DT D,

minute(d,n)

i=c+1

It was reported by [12 and 35] and it was shown that:

Equation 3 is obtained by the residual eigenvalues. The continuous solution of G lies under the c-1 space spanned by the first c-1 eigenvectors with the highest eigenvalues [12] DT D .

3.2. Clustering of spectral diagrams

Spectral graph clustering is popular because it is easy to program and offers a favorable balance between accuracy and computational complexity. Recently [11, 8] pointed out this point

Here, we introduced a weight matrix W for normalization purposes. Knockout

lutions. However, in our experimental results we encounterno problems of this type.

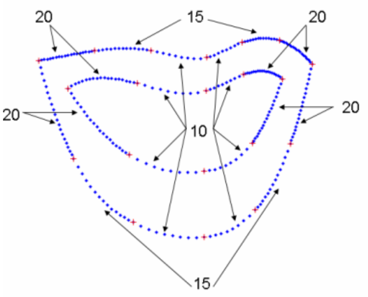


Figure 2. Number of samples in each segment.

Shape features alone cannot capture subtle differences in facial movements. For example, a figure can have two completely different poses (see bottom of Figure 3). To compensate for this effect, we include appearance features. Appearance features are extracted by a fixed geometric histogram, which was recently introduced in [13]. Since the histogram proposed in [13] is invariant to perspective transformations, the effect of registration on the appearance can be isolated.

−||si−Hij sj||2

Occurs at least p times (p is a user-specified criterion). In this section, we propose a simple but effective method to search for temporally coherent clusters.

ELIMINATION OF TIME REDUNDANCY

The first step is to automatically detect all neutral facial expressions (i.e. AU 0 in FACS) [5]. This is because these are usually the most common face 'clusters' and are useful in many recognition tasks. For example, the detection of subtle facial movements requires the calculation of the difference between the neutral image and the target image [24]. AU0 detection algorithm works as follows. First, calculate the normalized error between the shape/appearance at time t and time t 1 . Partition time temporally using a two-state hidden Markov model (HMM). Includes changes in appearance/shape. HMM transition probabilities are calculated using a logistic regression function. For state 0, which represents no change, the probability is given by 1, and similarly by 1 for other states. x is the normalized error, β, τ are parameters calculated from the error histogram. The standard Viterbi algorithm (dynamic programming) is performed to find the maximum posterior solution. It specifies a set of static facial expressions.

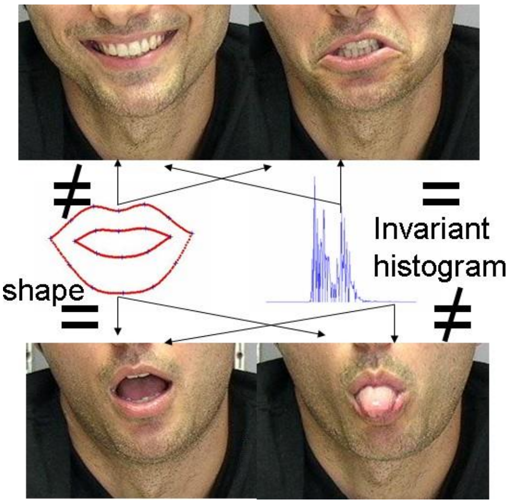


Figure 3. Features used for temporal segmentation.

Move more than 2 frames. This set includes AU 0

Shape. 3). The final inclination matrix is kij = e.

−||hi-hj||2

2σ2

Like other action units. The next step is to separate AU 0 from other AUs by performing spectral clustering.

a, where hi is the constant histogram of i sam-

ple (in a certain region) and σa is the standard deviation of the constant appearance.

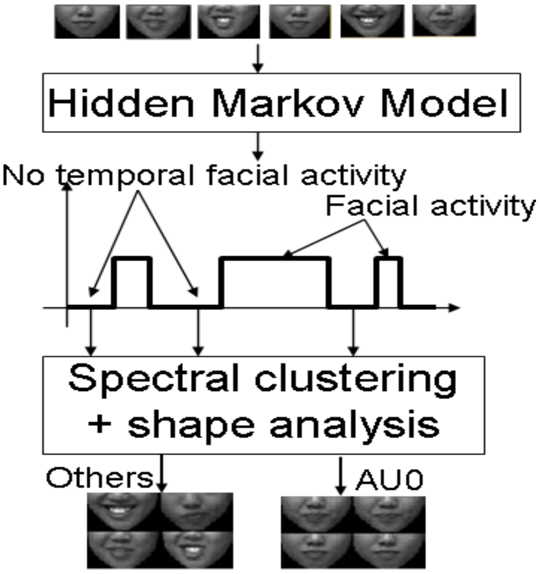


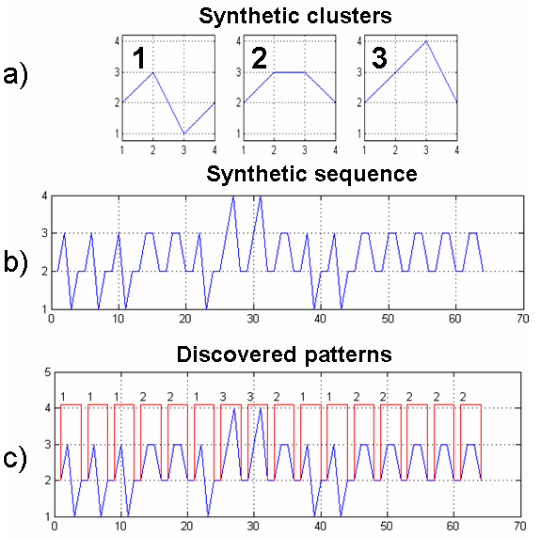
Figure 4. Process to automatically detect AU0.

The second step to discover temporal clusters is to achieve temporal invariance in the speed of facial movements. To achieve this goal, we first remove all consecutive frames belonging to a cluster. This method only saves cluster state changes. After this process is completed, the length of the video will be reduced to approximately 10-20% of its original size. Refer to the image. 8.a and 8.b.

TEMPORAL CORRELATION FOR FACIAL MOTION DETECTION

Once you've simplified the time representation of your video sequence, you're ready to find time patterns of varying lengths. The algorithm starts by choosing a long pattern (typically 8-9 consecutive clusters) as a pattern. Next, we calculate the normalized correlation between each pattern and sequence. All items with a normalized correlation of 1 (i.e., the same pattern as the template) are removed from the sequence. If your data is too noisy, you can apply a threshold smaller than 1. The algorithm then selects a smaller pattern (usually less than one cluster) and searches again for all instances of normalized correlation 1, continuing this step until all frames have been searched.

Figure 5 shows how the algorithm works on synthetic data containing three time clusters of length 4 (Figures 5.a and 5.b). The algorithm automatically identifies three time clusters.



EXPERIMENT

We evaluated this algorithm in two ways. First, we tested the ability to temporally segment facial movements and recognize rare movements. The latter was used to preprocess videos of spontaneous facial expressions for FACS annotation purposes.

TIME DIVISION OF ORAL EVENTS

In this experiment, we recorded a series of videos in which subjects spontaneously performed five different facial expressions: sad, tongue-in-cheek, talking, smiling, and neutral. We use a person-specific active appearance model [28, 9] to track non-rigid/rigid movements in the sequence (see Figure 6).

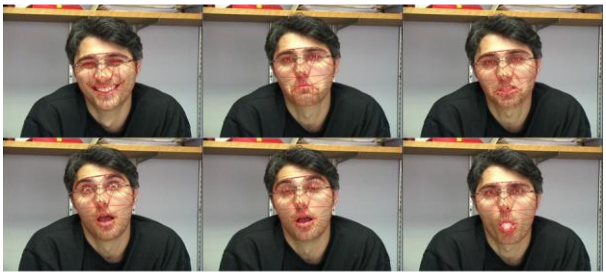


Figure 6. AAM tracking across several frames.

After tracking the video sequences using AAM, we use the algorithms proposed in Sections 4.1 and 4.2 to identify clusters (see Figure 7 for AU0) and remove temporal redundancy from the video sequences. By removing consecutive frames with the same cluster label, the length of the sequence is reduced to 20% of the original length (see Figures 8.a and 8.b). Then, the time segmentation algorithm detects the facial movements shown in 8.c. Note that there are some time periods that remain uncategorized. These windows correspond to moves that only last one frame, or unusual or rare moves.

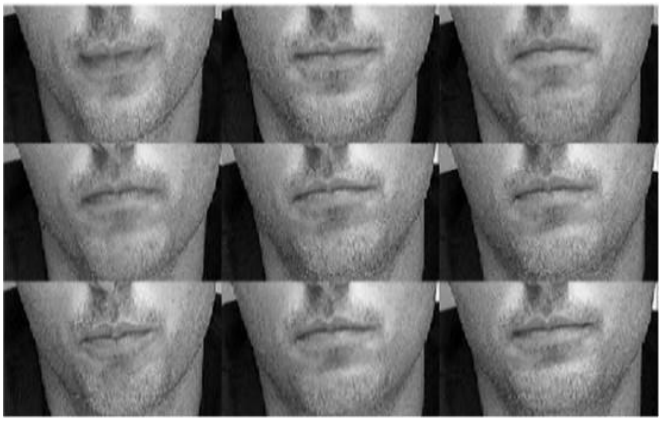


Figure 7. Examples of detected AU0.

The correctness of the clustering method was confirmed by visual inspection. Video results can be downloaded from www.cs.cmu.edu/˜ftorre/segfacialmotion.avi. Figure 9 shows a frame of the output video resulting from finding temporal clusters in a video sequence. Each frame of the video consists of three columns, where the first column displays the original image with a personalized AAM [28] model. The second column shows the prototype of each cluster identified by the algorithm. The third column shows all the facial movements in the video. per frame, cluster and time

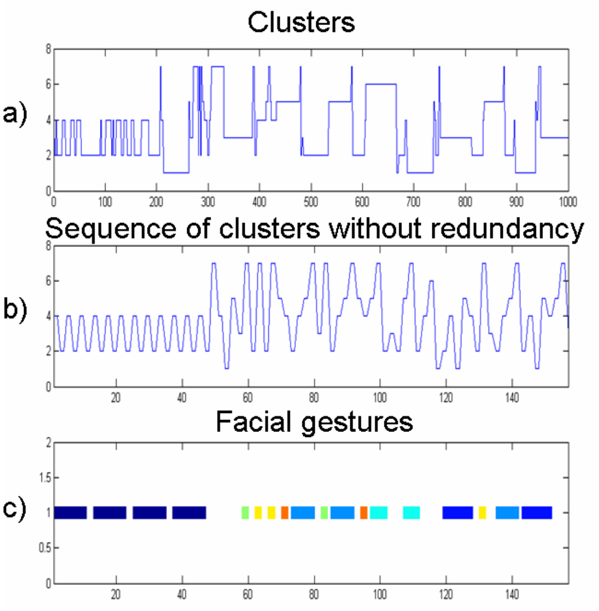


Figure 8. a) Original sequence of clusters. b) Sequence of clusterswith just the transitions. c) Discovered facial gestures.

gesture that corresponds to the image is highlighted.

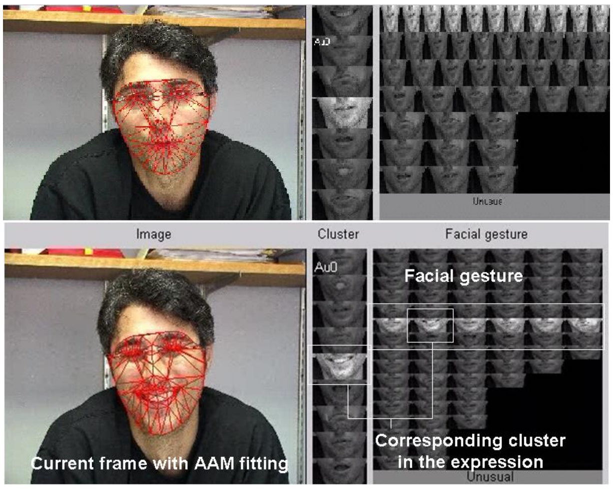


Figure 9. Frame of the output video.

Computer systems that improve the speed and reliability of manual FACS coding

FACS coding (Facial Action Coding System [14]) is the most advanced method in manual measurement of facial actions [5]. However, FACS coding is labor intensive and difficult to standardize. The goal of automatic FACS coding [6] is to eliminate the need for manual coding and achieve automatic recognition and analysis of facial movements. Completing the FACS coding required to train and test the algorithm was a speed limit. Manual FACS coding remains the same.

Thoughtful and slow. The speed, efficiency and quality control of FACS coding can be significantly improved by preprocessing the video streams for human coders using the time division proposed in this paper.

Current approaches to FACS coding

Currently, FACS encoders typically process video in one or more passes. When using the single-pass method, view the video and code the occurrence of all target action units in each frame. As they progress, they may need to attend to 20, 30, or even more units of action simultaneously. They are general oriented and need to monitor all possible units of action. Coding progress is slow (because all AUs must be considered) and quality is reduced because similar action units occurring in different frames cannot be checked together. Instances of a particular AU can be separated by long-term encoding of other AUs, which impairs the ability to visually recall past events. When the encoder proceeds in multiple passes, the quality improves because only a subset of AUs is encoded in each pass. In that case, the programmer becomes an expert and only looks for those few AUs, using the memory of what has already happened on the same topic. However, this process is inefficient because the encoder needs to display a potentially long video that does not contain the target AU. Increasing the time between encoding similar or related AUs impairs visual memory.

Subject	Accuracy	# of clusters	# of frames
19	68%	21	558

Table 1. Clustering Accuracy

The investigator said that he did not receive the check or not. We tracked the facial features of subject 19 using AAM [28, 9]. Remove AU0 using the steps described in Section 4. After preprocessing, we obtain 558 frames that are manually labeled into 21 AU units by a certified FACS coder. This manual FACS coding provides the ground truth for analysis.

From the shape and appearance data, compute a similarity matrix K containing the shape and appearance information and calculate the first 20 eigenvectors. Perform 50 K-means iterations in the embedding space and retain the solution with minimum error. To calculate the resulting accuracy for c clusters of cases using the ground truth, calculate a c × c confusion matrix C . where each entry cij is the number of instances of cluster i belonging to the class.

It is difficult to calculate accuracy using only j.

Since we do not know which cluster corresponds to which class, the confusion matrix C is used. The optimal way to solve the correspondence [20] is to calculate the following maximization problem.

Max tr(CP) |. The permutation matrix is (6).

Accuracy is obtained by dividing the result by the number of data points that are clustered. To solve equation 6, we use the classical Hungarian algorithm [20]. Table 1 shows the accuracy results. The clustering method is 68% consistent with manual annotation. This is comparable to interobserver agreement (70%) for manual coding.

It is interesting to note that the clustering results depend on the shape and appearance parameters, i.e. σs and σa. Figure 10 shows the accuracy as a function of these two parameters and shows that it is stable over a wide range of values.

CONCLUSION AND FUTURE CHALLENGES

In this paper, we presented a method for temporal segmentation of facial motion and demonstrated its usefulness in two new applications. This method is insensitive to geometric transformations, which is important in real-world environments where head movements are common. This method can be used to cluster similar facial movements, identify abnormal movements, and increase the reliability and efficiency of manual FACS annotations.

From

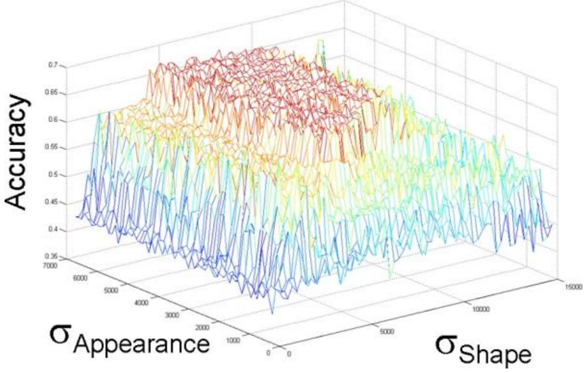


Figure 10. Accuracy variation versus σ_a and σ_s.

In addition, the lower surface of the face experiences asymmetric movements much more than other parts of the face due to the concentration of innervation on the opposite side [31]. To be useful, the system must include all areas of the face. In the present study, we extend the clustering to eye, midface, and eyebrow features.

Acknowledgments This research was supported in part by National Institute of Justice Award 2005-IJ-CX-K067 and National Institutes of Health Grant R01 MH 051435. Thanks to Iain Matthews for help with the AAM code and insightful comments on tracking. We would like to thank Thomas Simon and an anonymous reviewer for their helpful comments.

REFERENCES

[1] G.Adiv Determination of motion and three-dimensional structure of optical streams produced by multiple moving objects. IEEE Transactions on Pattern Analysis and Machine Intelligence, 7(4):384-401, July 1985.

[2] Z. Ambader, J. Schooler, and J.F. Cohn. Deciphering mysterious faces: The importance of facial dynamics in interpreting subtle facial expressions Psychological Science, 16:403-410, 2005.

[3] M. J. Black and Y. Yaqoub. Recognize facial expressions in image sequences using a local parametric model of image motion. International Journal of Computer Vision, 25(1): 23-48, 1997.

[4] J.F. Cohn. Automatic analysis of facial expression combination and time. In P. Ekman and E. Rosenberg, editors, What Faces Reveal: Basic and Applied On Expressions Using the Facial Action Coding System (FACS), Oxford University Collection of Emotional Sciences, 388–392, October 2005.

[5] J. F. Cohn, Z. Ambader and P. Ekman. Observer-based measurement of facial expressions using a facial action coding system. J. Coyne and J. Allen (eds.). The book of emotional arousal and evaluation. Oxford University Press Impact Science Collection. New York: Oxford, 2006.

[6] J. F. Cohn and T. Kanade. Using automatic facial image analysis to measure emotional expression. A book on emotional stimulation and evaluation. Emotional Science Series, Oxford University Press, New York, Oxford, 2007.

[7] J. F. Cohn and K. Schmidt. Timing of spontaneous facial gestures and movements