Intensity Score for Facial Actions Detection in Near-Frontal-View Face Sequences

Moi Hoon Yap
School of Computing, Mathematics, and Digital Technology
Manchester Metropolitan University
John Dalton Building, Chester Street
Manchester, M1 5GD, UK
M.Yap@mmu.ac.uk

Hassan Ugail
Centre for Visual Computing
University of Bradford
Bradford, BD7 1DP, UK
H.Ugail@bradford.ac.uk

Reyer Zwiggelaar
Department of Computer Science
Aberystwyth University
Aberystwyth, SY23 2AX, UK
rrz@aber.ac.uk

***Abstract*— this paper proposes a method to detect the facial Action Units (AUs) and introduce an automatic measurement in predicting the intensity scores of each AU in near-to-frontal face image sequences. To our knowledge, this is the first attempt in computer vision to automate the intensity scores of facial expression research. First, the facial feature points are detected by using Gabor feature based boosted classifiers and the movement of each point is tracked by optical flow. Then, we introduce a set of distance measurement for the feature points and analyze the distances by using Sequence Analysis. Further, we exploit the sequence partition to predict the possible temporal segments in the sequence. We found there is a relationship between the intensity scores and the partition threshold, which we automated the process of threshold selection in our work. We tested the proposed prototype with our in-house dataset and MMI database. Finally, we discuss the result, the possibilities of further research, and the next challenges for computer vision scientist in facial actions detection.**

Keywords- Facial Action; FACS; intensity score; facial expressions; temporal segments

I. INTRODUCTION

Human face is explained as the major communicative outputs and major sensory inputs [1]. According to Kong et al [2], the demand and research activities in machine recognition of human faces have increased significantly over the past 30 years. Such statement can be supported by the facts that a lot of past researches [3-5] and on-going researches [6] are trying to reveal the face information to provide clues/cues for entertainment, security and other technology development.

In psychology, it has a long history in facial behavioral analysis and measurement. The study of facial expressions has been conducted in last century [7, 8]. Facial Action Coding System (FACS) is less time consuming and cost saving compare to other facial behavioral measurement methods [9]. A few experiments have shown the popularity of FACS in research, for instance, Ekman et. al. [9] used FACS in predicting subjects' emotional experience while watching the films. Also, Kunz et. al. [10] has utilized FACS in pain assessment for dementia diagnostics.

In computer vision, the considerable progress has been made in automated facial expression analysis from digital video input within the past decade [4, 11]. Currently, there are a lot of researchers involved in the automatic recognition of AUs temporal segments. The existing approaches to facial expression analysis include geometric approach [12, 13], appearance-based approach [14], and Dynamic-Texture based approach [6]. However in the existing systems of automatic recognition of AUs temporal segments [6, 12, 13, 15], none of them make an attempt in recognising or predicting the intensity scores in the AUs. The novelties in this work are the implementation of sequence analysis in AUs detection and the introduction of an automatic method in intensity scores prediction. Intensity scores in facial expression are important in measuring the facial responses. Kunz et. al. [10] has demonstrated the contribution of the intensity scores in pain diagnostics for dementia research. They showed that the facial expression of pain, with intensity stimulation, has the potential to serve as alternative pain assessment tool in demented patients [10].

For clarity of presentation this paper has been divided into 5 sub-sections. Apart from this section which provides the reader an introduction to the problem domain and setting out the objectives of the proposed research, Section-2 introduces the background and research motivation. Section-3 proposes a methodology to detect the Action Units and intensity scores. Section-4 provides details of results and a comprehensive analysis. Finally section 5 provides the conclusions with an insight to possibilities of further research.

II. BACKGROUND AND MOTIVATION

The motivation for this work arises from a research project funded by UK Engineering Physical Sciences Research Council under the theme of the facial analysis for

This work is supported by EPSRC grant on "Facial Analysis for Real-Time Profiling" (EP/G004137/1).

real-time profiling project. The aim of the project is to provide a real-time dynamic passive profiling technique which will assist as a decision aid to Border Control Agencies, which has the potential to improve hit rates. The project is intended to combine and build on several research areas, which include multi-modality face and eye-movement tracking, eye-movement related to intent, dynamic thermal/visible face information related to intent, statistical shape and appearance modeling and face modeling and recognition.

Another motivation, like any other computer vision scientists, we attempt to automate the recognition and scoring of AUs. Face is the prominent visual object, and the expression is the basic mode of nonverbal communication among people [16]. It offers non-intrusive, perhaps the most natural way of authentication [2] – intuitive and does not stop user's activities. Research in measuring the face has been impeded by the problems of devising an adequate technique in distinguishing all possible visual facial movements. In psychology and human vision, Ekman et. al. introduced a way to recognize and score the AUs [17, 18]. This is not only a breakthrough in human vision and psychology, but it has motivated researchers from other domains (especially in computer vision) to work in this area. FACS is an extension of Facial Affect Scoring Technique (FAST) to distinguish all possible visually distinguishable facial movements [18]. It was built based on the neuroanatomy of facial behavior on the evolution of the nervous system and parallel elaborations of facial musculature [9]. There are three measurements in FACS: *Type*, *Intensity*, and *Timing*. The face is neutral when evidence of any specific AU is absent. *Type* refers to the action type, for instance, inner brow raise. *Intensity* refers to the magnitude of the appearance change. *Timing* refers to the duration of the movement. The five-point ordinal scale (A-B-C-D-E) is used to rank the present of the AU [18]:

A: trace of the action
B: slight evidence
C: marked or pronounced
D: severe or extreme
E: maximum evidence

The following section explains the methodology used in AUs detection and prediction of the intensity scoring.

III. METHODOLOGY

This section explains the methodology of our project, which includes facial feature point detection, facial feature point tracking, facial actions and analysis, and intensity scores measurement.

A. Facial Feature Point Detection

Among the facial feature point detection algorithms, Gabor feature based boosted classifiers is most robust at the time we write this paper. It is known that Gabor filters remove most of the variability in image in different illumination and contrast, it also robust against small shift and deformation [19]. In our work, we adopted the algorithm introduced by Vukadinovic and Pantic [19], which defines 20 facial points in images of neutral faces. This algorithm implemented Viola and Jones face detector [20] and divided the detected face into 20 regions of interest. Then, each of the regions is examined further to predict the location of the facial feature points. They used the individual feature patch templates to detect points in the relevant region of interest. The feature models are GentleBoost templates built from the gray level intensities and Gabor wavelet features [19]. Figure 1 illustrates the implementation of Gabor feature based boosted classifier on our in-house data. For further details of the Gabor Feature Based Boosted Classifiers in facial feature point detection, please refer to Vukadinovic and Pantic [19]. The executable version of the algorithm is freely available for research purpose from Pantic's personal website [21].

Figure 1. Facial feature points detection on our in-house database

B. Facial Point Tracking

In early tracking system, for e.g. in [22-24], feature matching was carried out from one frame to the next using optical flow computations. This has caused a problem which resulting in drifting errors accumulating over long image sequences. Face motion produces optical flow in image. The optical flow approach has the advantage of not requiring a feature detection stage of processing. Optical flow is the visible result of movement and is expressed in terms of velocity, it is a direct representation of facial action [24, 25]. Muscles actions can be directly observed in image sequence as optical flow, which is calculated by facial features and skin deformation [26]. Dense flow information was computed from images sequences of facial expression with Horn et al's [27] basic gradient algorithm, which is then reduced by taking the average length of directional components in the major directions of muscle contraction. Several muscle windows are located manually to define each muscle group using feature points as references. The muscle action derived from the muscle (group) model can be associated with several AUs of the FACS. In Mase's [25] experiment, he managed to identify 19 out of 22 test data in recognizing four expressions (happiness, anger, surprise and disgust) in motion. Due to the history and benefits of optical flow, we have adopted optical flow in our prototype.

C. *Facial Actions and Analysis*

Figure 2 shows the labeling of the facial feature points defined by Vukadinovic & Pantic [19]. They defined facial feature points as the corners of the eyes, corners of the eyebrows, corners and outer mid points of the lips, corners of the nostrils, tip of the nose, and the tip of the chin [19]. For instance, Point 19 shows the location of nose tip. The following explain the definition of each parameter used in facial action analysis:

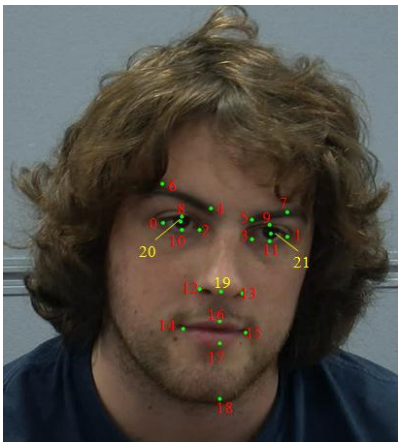

Figure 2. *Labeling of the facial feature points*

Eye Parameter:
Size of right eye (*ESizeR*) = Point 8 – Point 10
Size of left eye (*ESizeL*) = Point 9 – Point 11
Distance between right upper eye lid and nose tip (*EyeUpR*) = Point 8 – Point 19
Distance between left upper eye lid and nose tip (*EyeUpL*) = Point 9 – Point 19
Distance between right lower eye lid and nose tip (*EyeDownR*) = Point 10 – Point 19
Distance between left lower eye lid and nose tip (*EyeDownL*) = Point 11 – Point 19

Eye brows Parameter:
Eye brows distance (*BrowDist*) = Point 4 – Point 5
Distance between right inner brow and nose tip (*BInR*) = Point 4 – Point 19
Distance between left inner brow and nose tip (*BInL*) = Point 5 – Point 19
Distance between right outer brow and nose tip (*BOutR*) = Point 6 – Point 19
Distance between left outer brow and nose tip (*BOutL*) = Point7 – Point 19

Mouth Parameter:
Mouth size vertically (*MSizeV*) = Point 16 – Point 17
Mouth size horizontally (*MSizeH*) = Point 14 – Point 15
Distance between upper lip to nose tip (*MouthUp*) = Point 16 – Point 19
Distance between lower lip to nose tip (*MouthDown*) = Point 17 – Point 19
Distance between right lip corner to nose tip (*LipR*) = Point 14 – Point 19
Distance between left lip corner to nose tip (*LipL*) = Point 15 – Point 19

In facial action analysis, we assume each action consists of four temporal segments: Neutral, Onset, Apex, and Offset [18]. Neutral is the condition of expressionless face. Onset is the duration of the start frame of the action to apex. Apex is the duration of the greatest excursion of that action [18]. Offset is the duration of the end of Apex to Neutral. Figure 3 describes the temporal segments for the subject who expressed a happy expression. The first frame and the last frame of the sequence are the neutral state of the subject. The corner of the mouth start to pull from frame 15 to frame 25, where greatest pulling happens and the apex of this action is achieved. The subject holds the expression for 53 frames, at frame 78, she starts to resume to neutral state - the offset is from frame 78 to frame 88. The duration of temporal segments is very important in differentiate a blink from closed eyes, and possibly, in the future, to decide the micro-expressions.

TABLE I. PARAMETERS FOR FACE ACTION UNITS RECOGNITION

Action Unit (AU)	Name of AU	Parameters
AU0	Neutral	
AU1	Inner Brow Raiser	*BInR* AND *BInL* increase
AU2	Outer Brow Raiser	*BOutR* and *BOutL* increase
AU4	Brow Lowerer	*BrowDist*, *BInR*, *BInL*, *BOutR*, *BOutL* decrease
AU5	Upper Lid Raiser	*EsizeR* & *EsizeL* increase
AU6	Cheek Raiser and Lid Compressor	*EsizeR* & *EsizeL* decrease *EyeDownR* & *EyeDownL* increase
AU7	Lid Tighterner	*EsizeR* & *EsizeL* decrease *EyeUpR* & *EyeUpL* decrease
AU10	Upper Lip Raiser	*MouthUp* decreases *MSizeV* increases
AU12	Lip Corner Puller	*LipR* & *LipL* decrease *MSizeH* increases
AU15	Lip Corner Depressor	*LipR* & *LipL* increase *MSizeH* increases
AU25	Lips Part	*MSizeV* & *MouthDown* increase
AU45	Blink	Extreme changes in *EsizeR* & *EsizeL*
AU43	Eyes are closed completely	*EsizeR* & *EsizeL* ≈ 0 Only AU43E is available, no other intensity scores
AU46	Wink	Unilateral Blink (AU45) Extreme changes in *EsizeR* or *EsizeL*, but not both.

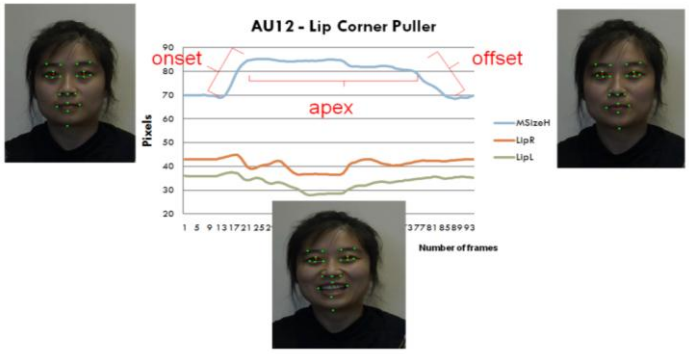

Figure 3. Analysis of temporal segments (onset, apex, and offset) in a sequence of frames

D. *Intensity Scores*

Due to the limitation of current research in computer vision for facial features detection/tracking, we reduced the five scales into three scales. We regroup the scoring into:

Slight: which consist of trace of an action and slight evidence, category A and B in FACS intensity scoring
Marked: marked or pronounced, category C in FACS intensity scoring
Extreme: which consist of severe or extreme, and maximum evidence, i.e. category D and E in FACS intensity scoring

Each of the movement tracking of the points will generate a sequence/series as show in figure 3. In order to automatically detect the hill or valley pattern in the sequence, we implement Sequence Partition [28] to split a sequence into equivalency classes. The categorical predictor is using a quadratic algorithm for splitting a set into one or more equivalency classes and the function returns the number of equivalency classes [29]. The comparison function is based on Euclidean distance.

In order to obtain an automatic intensity score, we make some assumptions: First, we expect the number of classes should be in between 3 and 15. Second, we predict the threshold's range is from 4 to 12. By default, the threshold is set to 7. The value of 7 represents the *Marked/Pronounced* of the facial action, the value from 4 to below 7 represent *Slight* of the facial action, and the value above 7 to 12 represents *Extreme* of the facial action. We programmed the prototype to automatically change the value of the threshold and stop when the number of classes is fall in between 3 and 15. Once we have the threshold fixed, we compare the classes to detect the hill and valley patterns.

We will further discuss and demonstrate the meaning of the intensity scores in Section 4.

IV. Result and Analysis

To explain the full functionality of our prototype, we illustrate the process by using a short clip from our in-house database. Figure 4 is a sequence of images from our in-house dataset, with the surprise expression.

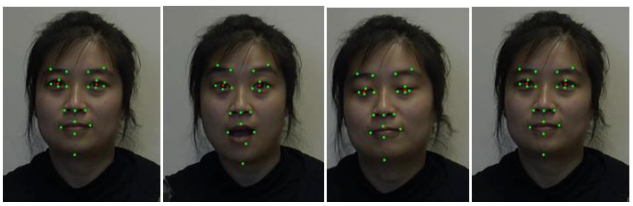

Figure 4. A sequence of images from a short clip of our in-house database

From this video, we observe *Slight*, *Marked*, and *Extreme* facial actions on the face. For instance, *Slight* for the distance between upper lip to nose tip (*MouthUp*), *Marked* for the distance between right inner brow and nose tip (*BInR*), and *Extreme* for the Mouth size vertically (*MSizeV*). Some of the screen shots of results of the AUs detection and automated intensity scoring are illustrated in figure 5. Figure 5(a) shows an increase distance between Upper Lip from nose with intensity 4, this can be interpreted as a *Slight* movement of the upper lip. Figure 5(b) shows an increase distance between the eye brows with intensity 7, we interpreted this as a *Marked* increase in the brows distance. Figure 5(c) shows the increment of the vertical size of the mouth with intensity 10. This can be interpreted as an *Extreme* lips part. In figure 5(d), the distance of the lower lip from the nose tip is increase with intensity 12. This figure shows an Extreme lower lip moving away from nose tip, adding the confidence of Extreme lip parts.

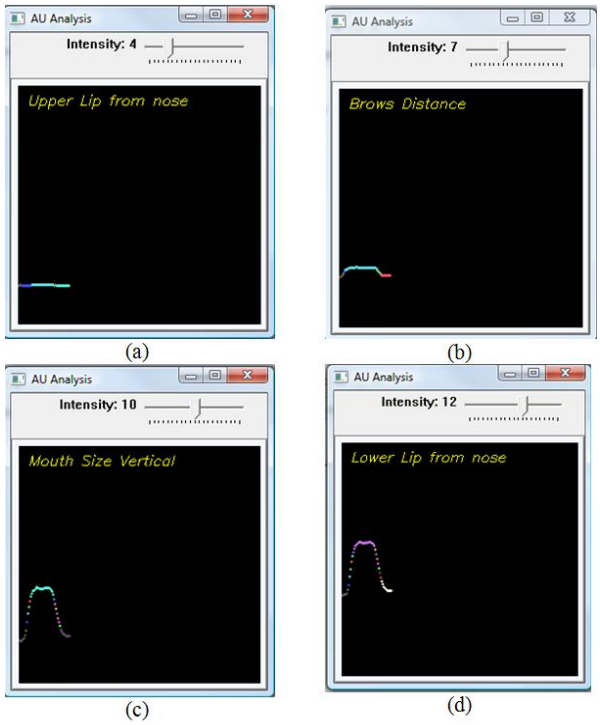

Figure 5. Screen Shots of results of the Aus detection and automated intensity scoring.

In addition, we also detect the wink and blink of the eyes by checking the size of the eyes. The distance of the eye lids is close to 0 if a wink or a blink is happened. For example, in this case, a blink is detected in this short clip at frame 43 for duration of 2 frames, as shown in figure 6. The prototype also generated a detailed report on the analysis at the end of the analysis, as in figure 6. From the report, we conclude that the short clip consists of *Slight* AU10, *Marked* AU1 and AU2, *Extreme* AU25, and AU46.

We have tested the prototype with MMI database [30]. From the database, we managed to detect AUs listed in Table I. The result is quite promising and the prediction is sound. In the case of miss detection/ failure in recognition, it is always caused by the error in computer vision algorithms, it is, detection errors and tracking errors.

```
Brows Distance:
getting further at frame 6 for the duration of 33 frames.
Intensity: 7
Brows Inner Right:
getting further at frame 5 for the duration of 33 frames.
Intensity: 7
Brows Outer Right:
getting further at frame 6 for the duration of 31 frames.
Intensity: 7
Brows Inner Left:
getting further at frame 6 for the duration of 32 frames.
Intensity: 7
Brows Outer Left:
getting further at frame 7 for the duration of 31 frames.
Intensity: 7
Lip Corner Right:
getting further at frame 11 for the duration of 24 frames.
Intensity: 7
Lip Corner Left:
getting further at frame 11 for the duration of 25 frames.
Intensity: 7
Mouth Vertical:
getting further at frame 12 for the duration of 24 frames.
Intensity: 10
Mouth Horizontal:
Upper Lip from nose:
getting closer together at frame 3 for the duration of 11 frames.
Intensity: 4
Lower Lip from nose:
getting further at frame 13 for the duration of 22 frames.
Intensity: 12
Left Eye: Blink at frame 43 for the duration of 2 frames
Right Eye: Blink at frame 43 for the duration of 2 frames

ACTION UNITS:
ACTION UNIT 1 - Inner Brow Raiser | Intensity Score: Marked
ACTION UNIT 2 - Outer Brow Raiser | Intensity Score: Marked
ACTION UNIT 25 - Lips part        | Intensity Score: Extreme
ACTION UNIT 10 - Upper Lip Raiser | Intensity Score: Slight
ACTION UNIT 46 - Blink
```

Figure 6. A Report on AUs recognition and intensity scores.

V. Conclusion

In this project, we take the first step in the attempt to predict the intensity scores of the Facial Action Units. Also, we introduce sequence analysis in automatically recognizing AUs and its' intensity scores. We believe that our work will motivate a lot of researchers in looking into the effort of intensity scoring automation and AUs recognition. The inaccuracy of the computer vision algorithms had a great impact on the facial action analysis. We will investigate in improving the facial feature tracking algorithm by replacing optical flow with particle filtering [31, 32]. The current prototype has a limitation: it only works on short clips (100 frames). In future, we will extend our work to long video clips analysis and recognize the AUs from the speech. In practical research, the subjects are normally facing an interview. It is a challenge for the FACS coder and researcher to recognize the AUs from the speech. On the other hand, we also plan to extend our work on the recognition of other AUs and the combination of AUs.

Acknowledgment

This work is supported by EPSRC grant on "Facial Analysis for Real-Time Profiling" (EP/G004137/1). The authors would like to thank Pantic et al for providing the MMI database.

References

[1] M. Pantic, and M. S. Bartlett, "Machine Analysis of Facial Expressions," *Face Recognition*, K. Delac and M. Grgic, eds., pp. 377-416, Vienna, Austria: I-Tech Education and Publishing, 2007.

[2] S. G. Kong, J. Heo, B. R. Abidi *et al.*, "Recent advances in visual and infrared face recognition - a review," *Computer Vision and Image Understanding, Elsevier Inc.*, vol. 37, pp. 103-135, 2005.

[3] R. Chellapa, C. L. Wilson, and S. Sirohey, "Human and Machine Recogntion of Faces: A Survey," in Proc. IEEE, 1995, pp. 705-740.

[4] B. Fasel, and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, vol. 36, no. 2003, pp. 259-275, 2003.

[5] M. H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting Faces in Images: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34-58, 2002.

[6] S. Koelstra, M. Pantic, and I. Patras, "A Dynamic texture based approach to recognition of facial actions and their temporal models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. In Press, 2010.

[7] C. Darwin, *The Expression of the emotions in man and animals*, London: J.Murray, 1872.

[8] P. Ekman, "Darwin and facial expression - a century of research in review," Academic Press New York and London, 1973, p.^pp. Pages.

[9] P. Ekman, W. V. Friesen, and J. C. Hager, "FACS Investigator's Guide," Research Nexus, a subsidiary of Network Information Research Corporation, 2002.

[10] M. Kunz, S. Scharmann, U. Hemmeter *et al.*, "The facial expression of pain in patients with dementia," *International Association for the Study of Pain, Elsevier.*, vol. 133, pp. 221-228, 2007.

[11] J. F. Cohn, and T. Kanade, "Use of Automated Facial Image Analysis for Measurement of Emotion Expression," *The handbook of emotion elicitation and assessment*, J. A. A. J. B. Coan, ed., p. 483, NewYork: Oxford: Oxford University Press Series in Affective Science, 2007.

[12] M. Pantic, and I. Patras, "Detecting Facial Actions and their Temporal Segments in Nearly Frontal-View Face image Sequences." pp. 3358-3363.

[13] M. Pantic, and I. Patras, "Dynamics of facial expressions - recognition of facial actions and their temporal segments from face profile image sequences," *IEEE Trans. Systems, Man and Cybernetics*, vol. 36, no. 2, pp. 433-449, 2006.

[14] M. Bartlett, G. Littlewort-Ford, M. Frank *et al.*, "Recognizing facial expression: machine learning and application to spontaneous behavior," in IEEE conference Comp. Vision and Pattern Recognition, 2005, pp. 568-573.

[15] M. Valstar, and M. Pantic, "Combined Support Vector Machines and Hidden Markov Models for Modeling Facial Action Temporal Dynamics," *Lecture Notes on Computer Science*, vol. 4796, pp. 118-127, 2007.

[16] J. C. Hager. "DataFace - Psychology, Appearance, and Behavior of Human Face," 05/05/2009, 2009; http://face-and-emotion.com/dataface/general/homepage.jsp.

[17] P. Ekman, and W. V. Friesen, “Facial Action Coding System: A Technique for the Measurement of Facial Movement,” *Consulting Psychologies Press* 1978.

[18] P. Ekman, W. V. Friesen, and J. C. Hager, "Facial Action Coding System - The Manual," Research Nexus division of Network Information Research Corporation, 2002.

[19] D. Vukadinovic, and M. Pantic, “Fully Automatic Facial Feature Point Detection Using Gabor Feature Based Boosted Classifiers,” in IEEE International Conference on Systems, Man and Cybernatics, Waikoloa, Hawaii, 2005, pp. 1692-1698.

[20] P. Viola, and M. J. Jones, “Rapid Object Detection using a Boosted Cascade of Simple features,” in CVPR, 2001, pp. 511-518.

[21] M. Pantic. "Maja Pantic," 17/08, 2010; http://www.doc.ic.ac.uk/~maja/.

[22] S. Basu, I. Essa, and P. A., *Motion regularization for model-based head tracking*, 362, MIT Media Laboratory Perceptual Computing Section, Cambridge, MA, 1996.

[23] I. Essa, T. Darnell, and P. A., "Tracking facial motion." pp. 36-42.

[24] K. Mase, and A. Pentland, “Lip reading by optical flow,” *IEICE of Japan*, vol. 6, pp. 796-803, 1990.

[25] K. Mase, and A. Pentland, “Recognition of facial expression from optical flow,” *IEICE, Trans. E*, vol. 74, no. 10, pp. 3474-3483, 1991.

[26] T. S. Huang, P. Burt, and K. Mase, "Computer Vision and Face Processing," *Final report to NSF of the planning workshop on facial expression understanding*, http://face-and-emotion.com/dataface/nsfrept/computer_vision.html, [26th May 2009, 1992].

[27] B. K. P. Horn, and B. G. Schunck, “Determining optical flow,” *Artificial Intelligence*, vol. 17, pp. 185-203, 1981.

[28] T. Hastie, R. Tibshirani, and J. Friedman, *Elements of Statistical Learning: Data Mining, Inference and Prediction*: Springer-Verlag, New York, 2001.

[29] G. K. Bradski, A., *Learning OpenCV: Computer Vision with the OpenCV Library*, First Edition ed., CA 95472, US.: O'Reilly Media, Inc., 2008.

[30] M. Pantic, M. F. Valstar, R. Rademaker *et al.*, "Web-based database for facial expression analysis."

[31] I. Patras, and M. Pantic, "Particle Filtering with Factorized Likelihoods for Tracking Facial Features."

[32] C. Su, Y. Zhuang, L. Huang *et al.*, "A Two-step Approach to Multiple Facial Feature Tracking: Temporal Particle Filter and Spatial Belief Propagation."