Machine Learning of Passengers Attention and Meditation during Autonomous Driving

Abe Hiroe[2], Luis Diago[1,2], and Ichiro Hagiwara[2]

[1] Interlocus Inc , Yokohama, Japan
[2] Meiji Institute for Autonomous Driving, Meiji University
Tokyo, Japan, {luis_diago,h_abe,ihagi}@meiji.ac.jp

Abstract. This work introduces a new framework to study the acceptance of an autonomous vehicle by multiple passengers. By using a *NeuroSky* biosensor, automatic annotation of facial expressions can be achieved to eliminates most of the challenges we faced due to the manual methods of annotating subjects emotional data. Two application scenarios are presented: simulated and autonomous cars. The experiments with simulated car passengers show a high correlation between facial parameters and emotional states measured from attention and meditation indexes. Prediction accuracy are at high levels of 73.3% and 61.9% for some subjects attention and meditation respectively.

Keywords: Autonomous vehicles, · human behavior learning, · facial expressions analysis · brain waves. .

1 Introduction

Driverless cars are seen as one of the key disruptors in the next technology revolution [1]. However, user acceptance needs to be dealt with in order for autonomous vehicles to be successfully introduced to the market [2]. Contrary to traditional vehicles where the acceptability of a model is basically determined by the acceptability of the vehicle by its driver, for autonomous cars it is necessary to analyze the acceptability for a group of passengers. Although "user acceptance" is an abstract term that has been redefined multiple times based on need and purpose, our goal in this research is to quantitatively measure emotional states of the passengers from the classification of their facial expressions and their correlation with attention indexes computed from their brain waves. Focusing on above goal, this work introduces a new framework to study passenger's acceptance of an autonomous vehicle. The framework (called "NeuroFaceLab") was introduced in our previous research [3] and it allowed analyzing the emotional states of the passengers of an autonomous vehicle. In this paper, we extend our previous framework to deal with the information coming from multiples passengers at the same time. The main problems in the application of the extended system are related with the identification of the sensors, their localization and the continuous recording of multiple passenger's emotional states.The main solutions developed to solve the above problems during the application of proposed framework are discussed in the paper.

In the following section 2, the architecture and two main scenarios of proposed framework are presented with the main problems faced during its application. In section 3, face parameters, their correlation with eSense values and the Long Short Term Memory (LSTM) network are presented with their related works. A comparison of the results of above methods within proposed framework is shown in section 4 and main conclusions and future works are presented in section 5.

2 NeuroFaceLab

Figure 1 shows the general architecture of proposed framework. The framework is a piconet network that allows one master device to interconnect with up to seven active slave devices using Bluetooth technology protocols. The application in the master device (the computer) has been developed using DirectShow library for rendering a video (Test.asf) and saving the input video coming from the video camera (WM ASF Writer) in combination with NeuroSky Inc software's tools for real time brain waves data acquisition. Serial communication using a COM port in the computer is used to acquire the *NeuroSky* data and save it to a TXT file for later analysis. All the information is synchronized in the application by using Source filter and Graph Manager in DirectShow. The framework also includes the possibility of video rendering during data acquisition.

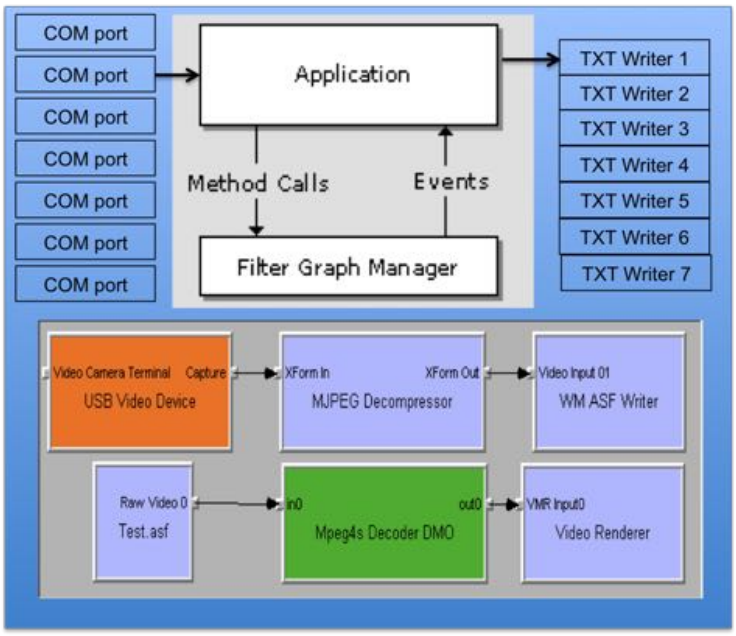


Fig. 1. Architecture of proposed framework

Fig. 2. Two application's scenarios: a) Simulator b) Autonomous car

Figure 2 shows two scenarios of application of proposed framework. In the first scenario (Fig 2a, simulator), a video of a driving in a simulator is shown and a video of the users is saved synchronously in combination with their brain waves during the user's observation of simulator's video. We have performed experiments in different uncontrolled environments where users have freedom of movement to observe the video in the simulator. The main problems related to the simulation environment were reported in our previous work [3]. They are the impossibility of detecting faces and tracking facial feature points robustly in uncontrolled environments. To solve above problems we have proposed a method that automatically corrects the brightness of the shadows in the face, and improves the detection accuracy of the characteristic points [4]. The method was extended for the processing of simulator videos. In this paper we focus on reporting the main problems related to scenario b (Autonomous car). The main problems arrive after we tested the proposed framework in a real experiment in Shodoshima Island, Japan during march 17-20/ 2019. The car shown in the picture has space for 8 passengers and was used following the specifications of the SAE level 3 in which the driver must be ready to intervene upon car's request. After the sensors are placed to the passengers, they decide to sit randomly inside the vehicle. The sensors also don't have a unique identifier to differentiate their transmitted information to the master node and this makes it difficult to identify the signals of each passenger. This problems are recognized in the literature [5] as the sensors identification and localization problem that has developed significant research interest among academia and research community. In addition to the aforementioned problems, the problem of continuous annotation of the emotional states of the passengers is added.

3 Annotation and prediction of eSense values

The availability of a large amount of labeled data is required for supervised Machine Learning (ML) approaches, especially research in face perception and emotion theory requires very large annotated databases of images of facial expressions with emotion. Data annotation is a time-consuming process posing major limitations to the development of human activity recognition systems [6].

Recently, real-time algorithm for the automatic annotation of a million facial expressions in the wild have been developed [7, 8]. However, the results of the challenge suggest that current computer vision and machine learning algorithms are unable to reliably detect 11 action units neither recognize 16 basic and compound emotion categories in images of facial expressions. Hence, the demands of real applications can drive the development of current algorithms.

Passengers' acceptance measurement is one of such applications. Passenger comfort experience during the autonomous driving essentially involves physical, physiological and psychological elements. In the context of autonomous cars it is essential that the developed ML models can explain the results of the inference made in an understandable way. As in out previous works [9], we focused on the interpretability of the computational models and try to extract rules explaining the emotional states of the passengers from the parameters computed from their facial expressions in order to find a correlation with the NeuroSky's eSense Levels obtained from the sensors.

3.1 Facial parameters

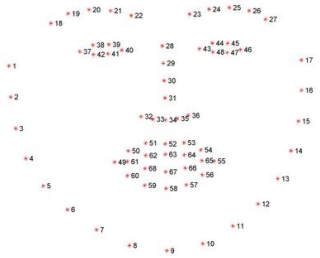
There are mainly two types of approaches for facial feature extraction [12–15]: geometric feature-based methods and appearance-based methods. The geometric facial features present the shape and location of facial components (including mouth, eyes, brows, nose, etc.) while in appearance-based methods, image filters, such as convolutional filters, are applied to either the whole face or specific region in a face image to extract a feature vector. When using one or the other approach? It is a question that remains in discussion among researchers. As our application focuses on the comprehensibility of nonlinear classifiers, facial expressions are represented by feature vectors composed by twenty parameters shown in Table 1. The table on the right shows areas $(p_1, \ldots, p_4)$ and distances $(p_5, \ldots, p_{20})$ computed from 68 feature points numbered on the left. Face-related parameters (e.g. eye blink) measured from the sensor are used for the identification of the sensors in scenario 2b.

3.2 NeuroSky's eSense Levels

eSense is a NeuroSky's proprietary algorithm for characterizing mental states. To calculate *eSense*, the NeuroSky thinkGear technology amplifies the raw brainwave signal and removes the ambient noise and muscle movement. The *eSense* algorithm is then applied to the remaining signal, resulting in the interpreted eSense meter values called Attention and Meditation (for short ATN and MED). ATN and MED from the subjects in the state of concentration, relax, fatigue and sleep have been analysed in previous research [10, 11]. The authors in [10] proposed a new method for detecting driving fatigue based on k-NN and the correlation coefficient r of subject's *eSense* values as follows:

$$r = \frac{\sum_{i=1}^{n}(x_i - x)(y_i - y)}{\sqrt{\sum_{i=1}^{n}(x_i - x)^2}\sqrt{\sum_{i=1}^{n}(y_i - y)^2}} \tag{1}$$

Table 1. Selection of 68 facial feature points and computation of face parameters: The column on the right shows areas $(p_1, \ldots, p_4)$, and distances $(p_5, \ldots, p_{20})$ computed from 68 feature points numbered on the left.

	**AREAS (A=Area_of_polygon):** $p_1(Right\ eye) = A(37-42)$ $p_2(Left\ eye) = A(43-48)$ $p_3(Mouth) = A(49-60)$ $p_4(Face) = A(18-27) + A(1,18,17,27) - A(1-17)$ **DISTANCES (D=Euclidean_distance):** $p_5(Vertical face) = \frac{1}{2}D(22,9) + \frac{1}{2}D(23,9)$ $p_6(Horizontal face) = D(17,1)$ $p_7(Eyes) = D(40,43)$ $p_8(Right\ eye - eyebrow) = \frac{1}{4}D(18,37) + \frac{1}{4}D(19,38) + \frac{1}{4}D(21,39) + \frac{1}{4}D(22,40)$ $p_9(Left\ eye - eyebrow) = \frac{1}{4}D(23,43) + \frac{1}{4}D(24,44) + \frac{1}{4}D(26,45) + \frac{1}{4}D(27,46)$ $p_{10}(Eyes - mouth) = height\_of\_triangle(37,46,58)$ $p_{11}(Vertical\ mouth) = D(52,58)$ $p_{12}(Horizontal\ mouth) = D(49,55)$ $p_{13}(Right\ eye - nose) = D(37,32)$ $p_{14}(Left\ eye - nose) = D(46,36)$ $p_{15}(Left) = D(22,49)$ $p_{16}(Right) = D(36,55)$ $p_{17}(Left) = \frac{1}{3}D(2,32) + \frac{1}{3}D(3,32) + \frac{1}{3}D(4,32)$ $p_{18}(Right) = \frac{1}{3}D(14,36) + \frac{1}{3}D(15,36) + \frac{1}{3}D(16,36)$ $p_{19}(Up) = \frac{1}{2}D(22,34) + \frac{1}{2}D(23,34)$ $p_{20}(Down) = \frac{1}{5}D(34,7) + \frac{1}{5}D(34,8) + \frac{1}{5}D(34,9) + \frac{1}{5}D(34,10) + \frac{1}{5}D(34,11)$

The x_i and y_i present the value of ATN and MED from the subjects left prefrontal lobe respectively at i moment, x and y present the average value of the *eSense* values respectively. Its sensitivity and specificity are at high levels of 68.31% and 90.43% respectively. Other authors [11] also measure and analyse *eSense* values to develop a driver warning system.The authors use a 5-point scale to classify the eSense values with the following ranges: 80-100 (high), 60-80 (little-high), 40-60 (neutral), 20-40 (low) and 1-20 (very-low). In case of the drivers' attention has been dispersed, it is aimed to provide the audio alerts to the drivers. As in the previous works [10, 11], we used r to analyze the relationship between facial expressions and eSense values. Instead of using r as a feature for learning, we use the 20 parameters shown in Table 1, apply a fussification to the eSense values and use r as a feature selector for the classifier.

3.3 Long Short Term Memory networks

Long Short Term Memory (LSTM) networks are a special kind of recurrent neural networks (RNN), capable of learning long-term dependencies [16]. In this work we used the Matlab 2018b implementation of the LSTM with the following parameters (MaxEpochs=10, MiniBatchSize=150, InitialLearnRate=0.01, SequenceLength=1000, GradientThreshold=1). The architecture of the LSTM was tested in two configurations: 20-100-3 and 1-100-3. In the first configuration the 20 parameters in Table 1 are used as input and in the second configuration only one input is used (the parameter with the higher correlation with eSense value). In both cases the number of hidden units is 100 and the number of outputs is 3. The adaptive moment estimation (ADAM) solver was used both cases.

4 Experimental Results

For a better use of space, only the results of 18 subjects who used the simulator in different environments are reported. As the stimulus is the same (video), the conditions are similar to those of the experiment carried out on the autonomous car. The results of the experiment with the autonomous car will be reported in future work. A video of 00:22:41 minutes with a resolution of 960x540 and a size of 360MB was shown to 18 subjects who participated in the experiments. The video includes three types of driving scenes: snowing in the morning (SM), sunny afternoon (SA) and dark night (DN). Fig.3 shows the results of face detection and tracking for 18 subjects who participated in the experiments. Several of the problems found in the autonomic car are presented in the simulation environment. As shown in the figure, the face was not detected in some subjects (e.g. row 3, columns 1 and 3) due to the movements of the subjects outside the visual angle of the camera. In other cases, even if the face is detected, the signals received from the sensors do not have a sufficient level (Poor Signal) so that the received ATN and MED levels are reliable.

Fig.4 shows an example of the time series for parameters $p_1, \ldots, p_{20}$ and eSense values (atn, med) of Subject 1(upper). The values have been normalized

Fig. 3. Results of face detection and tracking of 18 subjects

and rescaled for visualization. The full sequence has been divided into small sequences of 10-15 seconds each and used as train and validation sets. Odd sequences are used for training and even sequences for validation. Fig.4 also shows LSTM overtraining with the ATN data from Subject 1.

Fig.5 shows the correlation between facial parameters and brain signals for 12 subjects. In the correlation charts some subjects such as subject 5 exhibit a high positive correlation (0.3) with ATN values for all its parameters and others like subject 9 whose correlation with ATN is negative for all its parameters. Subject 5 also exhibits a uniform correlation (although not very high, 0.1) with the values of MED. Among all subjects, Subject 5 was the one who moved most during the simulation and changes in his facial parameters may be affected by abrupt changes in movement. Finally, Table 2 shows the learning results using an LSTM network for the classification of eSense values. The values of the 2nd (ATN) and the 4th (MED) column show the results obtained with the 20 parameters as features. On average, the prediction results of MED are superior to those of ATN. The 3rd and 5th columns show the results using a single feature (the one with the highest correlation coefficient r) for ATN and MED respectively. After using the correlation coefficient to select only 1 parameter as a feature for training and testing, as average it was possible to increase by 2.5 the accuracy of LSTM for the case of ATN values. For the case of MED values, a better strategy is needed for feature selection, although the percentage increases for some subjects (e.g. Subject 3 and 11).

5 Conclusion and Future Works

In this paper we presented the preliminary results in the development of NeuroFaceLab. This framework is a contribution trying to answer the question how potential users perceive the autonomous driving and if they accept this technology. We tried to quantify the impact of the autonomous driving on the mental conditions and emotional states of passengers by measuring their brain waves and classifying their facial expressions. Although the correlation between eSense values is very high (0.4703), the accuracy in the classification of Meditation index is higher than the same for Attention index if the 20 facial parameters proposed in the paper are used. Although prediction percent is at high levels for some subjects, feature selection strategies should be explored for the correct application of the proposed framework in the autonomous car scenario.

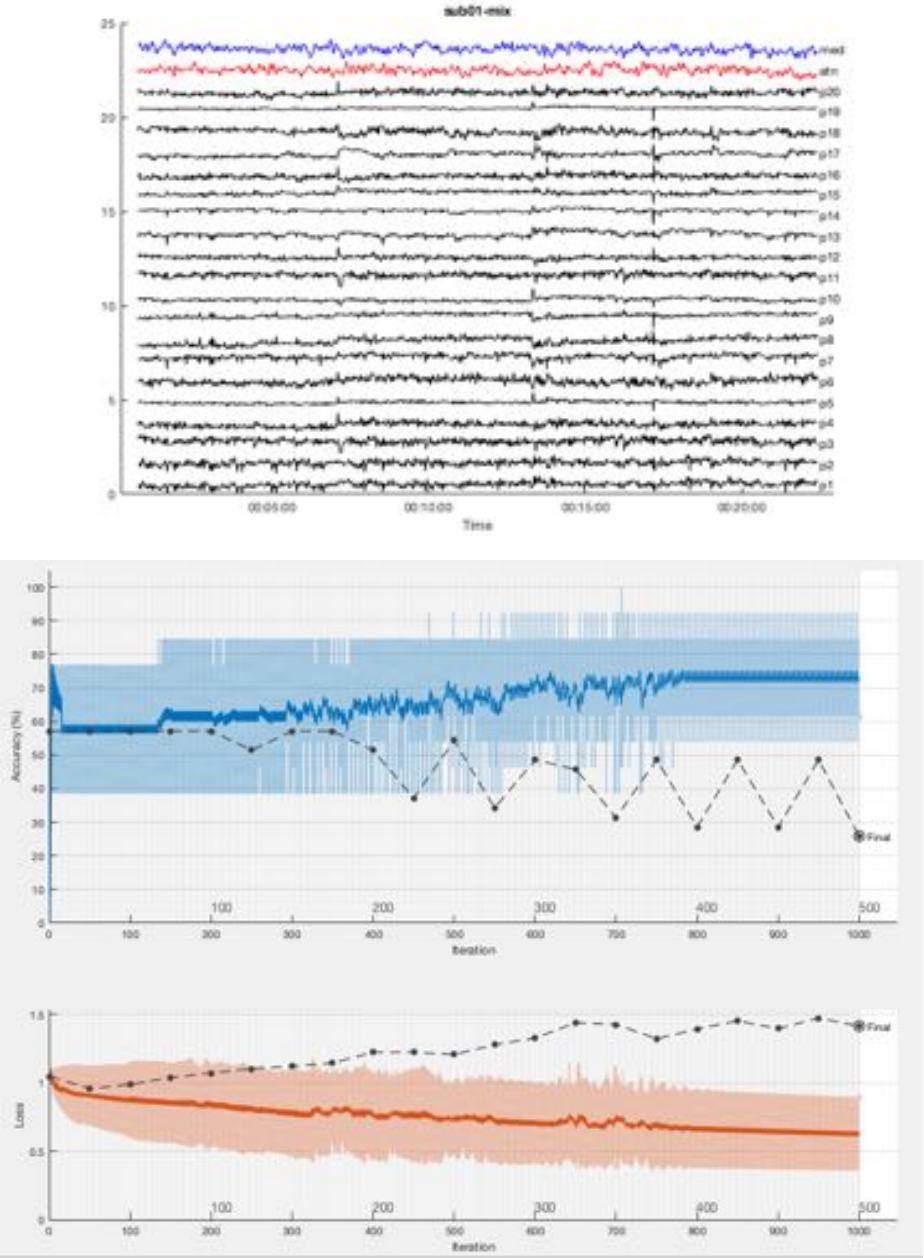


Fig. 4. Example of the time series for parameters $p_1, \ldots, p_{20}$, eSense values and a training process of LSTM with ATN data from Subject 1

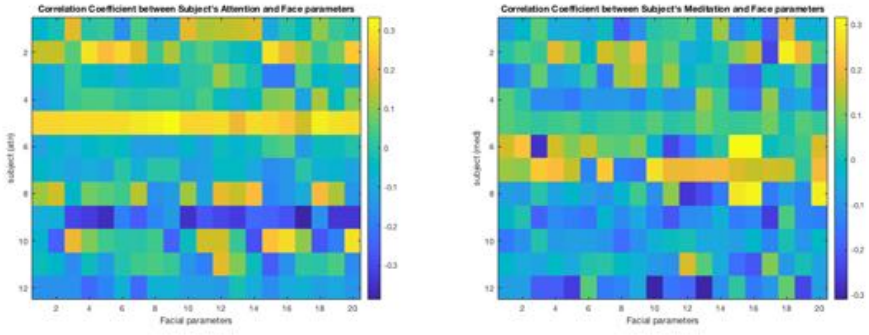


Fig. 5. Correlation between $p_1 ... p_{20}$ and eSense values (ATN -left and MED-right) for 12 subjects

Table 2. Validation accuracy (%) using LSTM network with 20 face parameters.

Subject	ATN	r	p	%	MED	r	p	%
1	25.71	0.1827	3	**45.71**	**45.71**	0.1984	17	42.86
2	**73.33**	0.2881	15	73.33	**56.67**	0.2891	18	40.00
3	**45.16**	0.2034	15	41.94	48.39	0.1991	20	**51.61**
4	30.95	0.1279	1	**33.33**	**61.90**	0.1744	20	40.58
5	**63.16**	0.3347	9	63.16	**52.63**	0.1005	17	36.84
6	*	0.1333	17	*	*	0.3169	16	*
7	*	0.1634	15	*	**50.00**	0.2754	10	50.00
8	47.06	0.2199	14	**52.94**	**47.06**	0.3067	16	47.06
9	**73.33**	0.3879	17	53.33	**60.00**	0.257	17	53.33
10	47.06	0.279	14	**52.94**	**41.18**	0.1695	8	35.29
11	26.67	0.2526	15	**46.67**	46.67	0.2045	16	**53.33**
12	**62.50**	0.2395	4	56.25	**56.25**	0.3107	13	37.50
Avr.	49.49	-	-	**51.96**	**51.97**	-	-	48.84

References

1. Kanwaldeep Kaur, Giselle Rampersad Trust in driverless cars: Investigating key factors influencing the adoption of driverless cars Journal of Engineering and Technology Management (2018), Vol. 48, April-June 2018, Pages 87-96
2. P. Wintersberger, A. Riener and Anna-Katharina Frison Automated Driving System, Male, or Female Driver: Who'd You Prefer? Comparative Analysis of Passengers Mental Conditions, Emotional States & Qualitative Feedback in Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive V ehicular Applications, 2016, Pages 51-58
3. Diago, L., Yang, Y., Abe, H., Hagiwara, I.: NeuroFaceLab : A new framework for passengers analysis in autonomous driving. In: Proceedings of the 31st International Computational Mechanics Symposium - CMD2018. The Japan Society of Mechanical Engineers, Comptational Mechanics Division, Tokushima, Japan (2018)
4. Toshiki Kanbayashi, Luis Diago, Tetsuko Kitaoka, Ichiro Hagiwara, *Examination of Correction Method to Shadow in Face Image for Iyashi Expression Recognition System*, *The Journal of the Institute of Image Electronics Engineers of Japan*, 2012, Volume 41, Issue 1, Pages 28-35,
5. Anup Kumar and Takuro Sato, *Localization in Wireless Sensor Networks: A Survey on Algorithms, Measurement Techniques, Applications and Challenges J. Sens. Actuator Netw.* , 2017, 6, 24; doi:10.3390/jsan6040024
6. Federico Cruciani , Ian Cleland , Chris Nugent, Paul McCullagh, Kre Synnes and Josef Hallberg *Automatic Annotation for Human Activity Recognition in Free Living Using a Smartphone*, Sensors 2018, 18, 2203; doi:10.3390/s18072203
7. C. F. Benitez-Quiroz, R. Srinivasan and A. M. Martinez, *EmotioNet: An Accurate, Real-Time Algorithm for the Automatic Annotation of a Million Facial Expressions in the Wild* 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 5562-5570. doi: 10.1109/CVPR.2016.600
8. Benitez-Quiroz, Carlos F. et al. *EmotioNet Challenge: Recognition of facial expressions of emotion in the wild* , CoRR abs/1703.01210 (2017)
9. Luis Diago, Tetsuko Kitaoka, Ichiro Hagiwara and Toshiki Kambayashi *Neuro-Fuzzy Quantification of Personal Perceptions of Facial Images Based on a Limited Data Set IEEE Transactions on Neural Networks* 22(12):2422-34
10. Jian He, Dongdong Liu, Zhijiang Wan, Chen Hu A noninvasive real-time driving fatigue detection technology based on left prefrontal Attention and Meditation EEG 2014 International Conference on Multisensor Fusion and Information Integration for Intelligent Systems (MFI).
11. Nilay Yildirim and Asaf Varol Warning System for Drivers according to Attention and Meditation Status Using Brain Computer Interface, IJAECS, 2016.
12. Maja Pantic and Leon J. M. Rothkrantz. *Automatic analysis of facial expressions: The state of the art IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, 2000.
13. B. Fasel and J. Luettin. Automatic facial expression analysis: A survey *Pattern Recognition*, 36(1):259–275, 2003.
14. Y Tian, T Kanade, and J Cohn. *Facial expression analysis*, chapter 11. Springer. New York, 2005.
15. M. Pantic and MS. Bartlett, *Machine Analysis of Facial Expressions*, in (Chapter 20) , Edited by D. A. White and D. A. Sofge, I-Tech, Viena, Austria,2007.
16. Sepp Hochreiter; Jrgen Schmidhuber *Long short-term memory. Neural Computation.* 9 (8): 1735?1780. doi:10.1162/neco.1997.9.8.1735. PMID 9377276.