Automatic Analysis of Facial Action

Krist Fama

28 February, 2018

Abstract

Face expressions recognition has recently received a significant amount of attention. Over the past 30 years a lot of paper were written describing different features used in face expressions recognition system. This paper provides an overview of all these features and describes some of them in detail. Action Units (AU) are defined and summarized as part of the Facial Action Coding System (FACS). They are significant components regarding face description. There are three underlying motivations for me to write this paper. First of all, I want to give a short review of how the subject of face expressions recognition came to be and where this topic it is headed. The second motivation is to reveal the differences between various approaches used in each step of the face expressions recognition system and finally to show the challenges and opportunities in this field.

1 Introduction

Scientific work on face expressions can be traced back to 1862, with the work of the french researcher Duchene [12]. He studied the electro-stimulation of individual facial muscles responsible for producing facial expressions and left a very detailed database with black and white photographs. In 1872, Charles Darwin published "The Expression of the emotions in Man and Animals"[3], in which the importance of facial expressions for communication and description was explored.

Nowadays, facial expressions serve as primary nonverbal means for human beings to regulate their interactions [21]. They clarify and emphasize what is being said and help us show our emotions. Paul Ekman suggested the six basic emotions: anger, fear, disgust, happiness, sadness and surprise to describe the facial expressions. A lot of possible facial expressions can be described with these six emotions or the combination of them [20]. The facial expression measurement in this case is a message judgment, which means that the facial expressions are only measured by the facial display, such as being sad or happy. An objective method to measure the facial expressions is the sign judgment. In this method, the physical signals such as depressed lips or raised cheeks are studied. The most common descriptor in sign-judgment are those described by the FACS, originally developed by Ekman and Friesen in 1978 [22] and revised in 2002. Specified are 32 atomic facial muscle actions, referred to as AU and 14 additional Action Descriptors (ADs). This paper focuses only on the AUs.

Most of the research done by psychologists or neuroscientists is done using FACS for different aspects of facial expression analysis. Various research is also done in Otto-Friedrich-Universität in Bamberg at the Cognitive Systems Department as part of student projects in recent years. Faces from different people, while being in pain, are video-recorded and the muscular movements are saved in a database, each with a belonging facial action code.

Figure 1: Expressions of grief from "The expression of the emotions in man and animals"-London, 1872 [3].

This work provides an overview of different features, used in a face expressions recognition system. The paper is structured into seven sections. The following section presents a brief review of FACS with all the corresponding AUs. Section 3 describes shortly a facial action recognition system. Section 3 provides a summary of pre-processing of the data, with all the components needed, in order to detect, localize and register the facial expression of a human being. Section 4 contains a detailed review of feature extraction with the appearance, geometric, motion and other features needed to make the image appear as clear as possible. Section 5 summarizes the state of the machine analyses of facial actions and analysis the individuals AUs, temporal consistency and spatial relations. Finally, section 7 discusses the challenges and the opportunities in future works and gives a conclusion of this paper.

2 Facial action coding system

The FACS is a comprehensive, anatomically based system for measuring all visually discernible facial movements. FACS describes all visually distinguishable facial activity on the basis of 44 unique AUs, as well as several categories of head and eye position and movement.

Every possible facial expression can be objectively described as a combination of AUs. A list with the AUs coded in FACS and the muscle groups involved in each action are shown in Fig 2. and Fig.3 [23].

AU number	Descriptor	Muscular Basis
1.	Inner Brow Raiser	Frontalis, Pars Medialis
2.	Outer Brow Raiser	Frontalis, Pars Lateralis
4.	Brow Lowerer	Depressor Glabellae, Depressor Supercilli; Corrugator
5.	Upper Lid Raiser	Levator Palpebrae Superioris
6.	Cheek Raiser	Orbicularis Oculi, Pars Orbitalis
7.	Lid Tightener	Orbicularis Oculi, Pars Palebralis
9.	Nose Wrinkler	Levator Labii Superioris, Alaeque Nasi
10.	Upper Lip Raiser	Levator Labii Superioris, Caput Infraorbitalis
11.	Nasolabial Fold Deepener	Zygomatic Minor
12.	Lip Corner Puller	Zygomatic Major
13.	Cheek Puffer	Caninus
14.	Dimpler	Buccinator
15.	Lip Corner Depressor	Triangularis
16.	Lower Lip Depressor	Depressor Labii
17.	Chin Raiser	Mentalis
18.	Lip Puckerer	Incisivii Labii Superioris; Incisivii Labii Inferioris
20.	Lip Stretcher	Risorius
22.	Lip Funneler	Orbicularis Oris
23.	Lip Tightener	Orbicularis Oris
24.	Lip Pressor	Orbicularis Oris
25.	Lips Part	Depressor Labii, or Relaxation of Mentalis or Orbicularis Oris
26.	Jaw Drop	Masetter; Temporal and Internal Pterygoid Relaxed
27.	Mouth Stretch	Pterygoids; Digastric
28.	Lip Suck	Orbicularis Oris

Figure 2: Single AUs in FACS, as described in [23].

AUs can occur either alone or in combination. When AUs occur in combination they can be additive, which means that the combination does not change the appearance of the constituent AUs. If they are non-additive, the appearance of the constituents does change. Although the number of atomic action units is relatively small, more than 7,000 different AU combinations have been observed. FACS provides the descriptive power necessary to describe the details of facial expression [29]. With facial action unit coding is possible to dissociate between the three following facial expression categories:

- Macroexpressions
  Typically last between 0,5-4 seconds, occur in daily interactions and are generally obvious.

- Microexpressions
  Last less than half a second, occur when trying to consciously or unconsciously conceal or repress the current emotional state. Microexpressions are much harder to be detected than macroexpressions.

- Subtle expressions
  These expressions are associated with the intensity and depth of the underlying emotion. The intensity of these facial actions constantly varies.

AU number	FACS name
8.	Lips Toward Each Other
19.	Tongue Out
21.	Neck Tightener
29.	Jaw Thrust
30.	Jaw Sideways
31.	Jaw Clencher
32.	Lip Bite
33.	Blow
34.	Puff
35.	Cheek Suck
36.	Tongue Bulge
37.	Lip Wipe
38.	Nostril Dilator
39.	Nostril Compressor
43.	Eyes Closure
45.	Blink
46.	Wink

Figure 3: More grossly defined AUs in the Facial Action Coding System[23].

Facial expression can also be coded in terms of "events". An event is the AU-based description of each facial expression, which may consist of a single AU or many AUs contracted as a single expression [23].

3 Facial action recognition system

A facial action recognition system is a complex combination of three main steps which include pre-processing of the data, feature extraction of the data after pre–processing and than the machine analyzes the facial action. In the Fig.4 is shown a facial action recognition system and the categories of each main step are listed. In the following sections each of these steps and their different approaches are described.

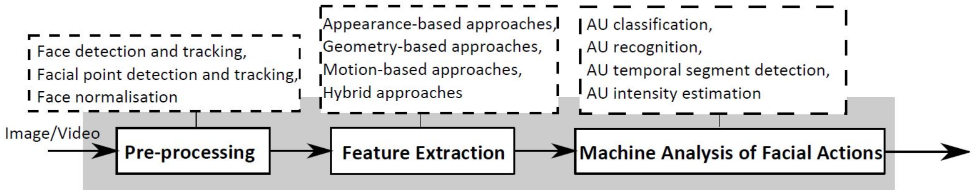

Figure 4: Configuration of a generic facial action recognition system [1].

4 Pre-processing

The first step done in a facial action recognition system is pre-processing, as seen also in Fig.4. The aim of pre-processing is to align faces into a common reference frame, so that the features extracted from each face correspond to the same semantic locations [1]. In this step, rigid head motions and anthropomorphic variations between people, are removed. Pre-processing includes three components: the face detection, facial landmark localization and face normalization. The following subsections describe these three components briefly.

4.1 Face detection

The goal of face detection is to determine whether there are any faces in the image and, if present, return the image location and extent of each face. It is not trivial to detect faces. Some known face detection challenges are pose, occlusion, expression and image conditions.

One of the most widely known face detector is the Viola and Jones algorithm. This simple and efficient classifier is built using the AdaBoost learning algorithm and three different geometric features to detect the face [25]. Pre-trained models of this algorithm can be found in OpenCV [1] or Matlab [2] and this is one of the reasons why this algorithm the reference face detection algorithm is. Another open-source face detector is the one provided with the Dlib [3]. The pre-trained facial landmark detector inside the dlib is used to estimate the location of 68 (x, y)-coordinates that map to facial structures on the face and is also widely used. Recent works have successfully adapted the deformable parts model (DPM) to perform face detection. An improved detection robustness and localization

[1] http://opencv.org/
[2] https://de.mathworks.com/products/matlab.html
[3] http://dlib.net/

accuracy was achieved at a higher computational cost[1]. Other ideas involve the use of deep learning algorithms in face detection.

4.2 Facial landmark localization

Face landmarking is defined as the detection and localization of certain characteristic points on the face. Commonly used landmarks are among others the eye corners, chin, center of the bottom lip, the nose tip and the eyebrows. Taken together, they define the shape of the face.

Facial landmark localizations are categorized in generative and discriminative models. Generative models are defined with the Active Appearance Model (AAM) proposed by Cootes et al. [6]. This model finds the optimal parameters for both: the face shape and the face appearance. These optimally reconstruct the current face. The shape is optimized by minimizing the texture representation error. Once the model is fitted on an input image, the optimized model parameters are used for facial expressions recognition [27]. Discriminative models typically represent the face appearance by considering small patches around the facial landmarks. Active Shape Model (ASM), also proposed from Cootes et al. [6], is part of discriminative models and has been very popular due to its earlier success. This local appearance model allows an efficient search to be conducted in order to find the best candidate point for each landmark.

Two other categories related to facial landmark localizations are the primary and secondary landmarks. The directly detected landmarks referred to as primary or fiducial ones, and they play a determining role in facial identity. The landmarks in the secondary category such es nostrils, chin, cheek contours, eyelids often present scant image evidence. These landmarks are also known as ancillary ones. Additional information about these categories can be found in [19].

4.3 Face registration

In this step each face is registered to a common pre-defined reference coordinate system. Depending on the output of the registration process, registration strategies are categorized as whole face, point and part registration. These three categories are described in the following text.

- Whole face registration
  The region of interest for most systems is the whole face. The Registration of the whole face can be made with rigid or non-rigid registration techniques. Rigid Registration is generally performed by detecting facial landmark and using their location to compute a global transformation that maps an input face to a prototypical face. Non-rigid approaches enable registration locally. Registration errors due to facial activity can be prevented. For both these types of registration AAM are used but also other techniques such as Robust FFT for rigid ones or SIFT-flow for non-rigid ones. In [26], [13] and [4] these techniques are briefly described.

- Point registration
  This type of registration is needed for shape representation, for which registration involves the localization of fiducial points also known as secondary landmarks. For Point registration AAM and facial feature detectors are used.

- Part Registration
  In this type of registration the face is processed in terms of parts (e.g eyes, mouth, etc). The size, number and location of parts to be registered may vary. In order to make this registration, similar to the other registration methods, the AAM is used. The parts are typically localized as fixed size patches around detected landmarks [11].

All three categories have their advantages and disadvantages in different problems. While some representations are coupled with a certain type of registration only, others are also used in combination to achieve the best result.

5 Feature extraction

Feature Extraction means converting image pixel data into a higher level of representation. It reduces the dimensionality of the input space and it minimizes the variance in the data caused by unwanted conditions such as lightening, blur error, head pose and others. Image cropping, resizing and changing the brightness are some steps of the feature extraction which can be made after the data comes from detection. The extraction features are categorized in appearance, geometric, motions and deeply learned. These features with some of their subcategories are briefly described in the following subsections [1].

5.1 Appearance features

Appearance features describe the texture and color of the face and are the most commonly used ones. They can be used with any given AU and the image is processed as two dimensional patterns.

The concept of "feature" in this approach is different from simple facial features such as eyes and mouth. Any extracted characteristic from the image is referred to as a feature. The appearance based approach keeps important informations of images and rejects the redundant information. Methods such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are used in order to extract the feature vector.

The main purpose is to reduce the large dimensionality of observed variable to the smaller intrinsic dimensionality of independent variable without losing much information. Appearance features can be characterized in terms of the representation strategy, in holistic features and local features [1].

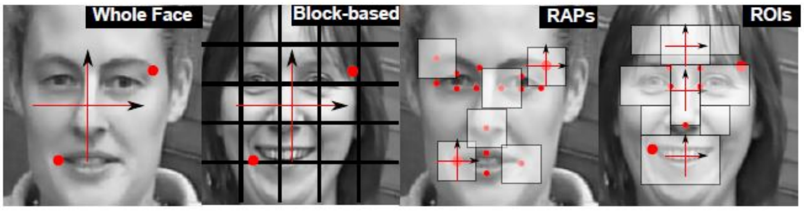

Figure 5: Different ways to apply appearance descriptors. Left to right: whole face, block-based, RAPs and ROI defined by points. The first two representations are holistic, second two are local [1].

- Holistic features
  In this approach the whole face region is taken into account as input data. Holistic features are defined as those that extract information according to a coordinate system relative to the entire face. These features require a preprocessing procedure to normalize the face image variations in pose and scale. This is a challenging task since it depends on the accurate detection of at least two landmarks from the face image [26]. Methods like fisher face, Support Vector Machine (SVM) and Hidden Markov Model (HMM) fall under this approach. The first two representations in the Fig.5 are holistic feature based approaches.

- Local features
  These features consider locations relative to a coordinate system defined by inner-facial features such as facial components or facial points. Some local approaches consider small patches centered around each of the facial landmarks or a smaller group of them. After this, for each of the patches a feature descriptor is applied. The resulting descriptors from all the patches are then used to build the final feature vector [15]. The third and fourth representation in Fig.5 present Region Around Points (RAP) and Region Of Interest (ROI), based on the local features.

In contrast to holistic methods, some local feature-based approaches for face expression recognition [8], [9], [10], are more robust to image variations in pose and scale. Furthermore, different from the holistic approaches, the local features have the face normalization as an integrated part of the approach. Other advantages are seen in illumination changes and non-frontal head poses, which can easily be locally approximated. Although some local feature based approaches achieve better recognition performances than holistic approaches [8], their computational burden is much heavier. Nevertheless local representations are nowadays generally more preferred.

5.2 Geometric features

In this technique, features are extracted using the size and the relative position of important image components. First the direction and edges of most significant components are detected and then the feature vectors are built.

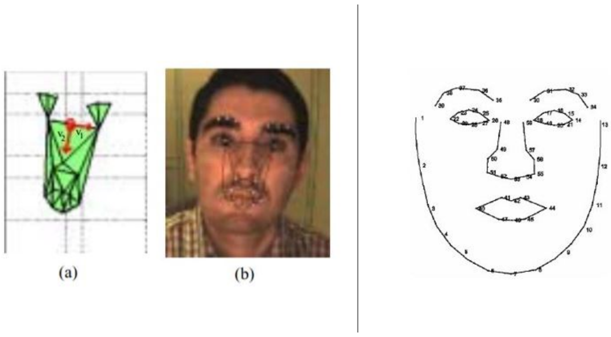

Figure 6: Left: (a) The N = 19 point mesh and the two vectors used to align the mesh across subjects. (b) The mesh projected on the image of the face [14]. Right: Facial landmark (58 points) [7].

Typical examples of geometric-feature-based methods are those of Gokturk et al.[14] who used 19 point face mesh or of Chang et al. [7] who used a shape model defined by 58 facial landmarks. Fig.6 represents these two geometric feature based approaches. The geometric features are easy to register and different from appearance features, these features do not show problems when it comes to lightening conditions. Yet, it seems that using both geometric and appearance features might be the best choice in case of certain facial expressions.

6 Machine analysis of facial actions

When using various machine learning techniques to deal with AUs, different problems are distinguished: AU detection, AU intensity estimation, AU temporal segment detection and AU classification [1]. These problems are characterized by temporal and spatial correlations. Temporal correlations are related to the nature of the input data. The spatial ones refer to the fact that some AUs may co-occur. A specific overview of all the different techniques used for AU analysis are described in [5]. In the following section these four problem are highlighted and different techniques are described.

6.1 AU classification

AU problems are characterizes by important temporal and spatial correlations. Temporal correlations are related to the constraints resulting from the temporal nature of data. In contrast, spatial correlations refer to the fact that some AUs tend to co-occur. These correlations are mostly captured with different techniques in frame-based approaches.

In the following some approaches using Spatial and Spatial-temporal approaches are described.

- Spatial classification approaches
  Neural network is widely used in AU classification with spatial correlations. However, training a neural network is difficult for spontaneous facial behavior, in which over thousands of AU combinations have been observed. A lot of researches are made using SVM in order to identify AUs. The best recognition performance till now is obtained through SVM classification on a set of Gabor wavelet coefficients selected by Adaboost. The computational complexity of SVMs however, is still considerable [32]. One limitation of spatial correlations is the fact that they only attempt to classify each AU independently, ignoring the semantic relationship among AUs.

- Spatial-temporal classification approaches
  Spatial and Temporal correlations can also be combined in the so called Spatial-Temporal relations. Capturing both has the potential for further performance benefits. Some works which combine these two types of correlations using HMM, Bayesian Networks (BN) and SVM are [30], [31] and [32]. In the attempts done using HMM, the large number of HMMs required to identify a great number of potential AU combinations prevents it from real-time approaches. Researches with BN showed that this methods cannot handle complex relationships between facial features, as well as temporal changes.

Although a lot of work is done in this direction, still it is not easy to deal with the potential thousands of AU combinations.

6.2 AU detection

Different databases which serve as input data are composed of video sequences. In these videos the target AU may occur at any time or may not occur at all. Two approaches in detecting AUs are categorized in frame-level approaches and segment-level approaches.

Frame-level Approaches perform inference at each frame of the sequence, assigning one of the target labels to each of them. Segment-based approaches focus instead on localizing events as a whole, taking an input of a representation of a spatio-temporal data segment[1]. In the frame level approaches inconsistency is seen sometimes where single frames, within a sequence, have a different target AU as all other group frames. A good performance is shown, if this approach is combined with temporal consistency information, typically done through the use of a graphical model. Common binary classifiers applied in this approach are Artificial Neural Networks (ANN), SVM and Boosting techniques. SVM are nowadays the most popular choice when detecting AUs. Furthermore, the segment-based approaches are not really used since they are really complicated to implement and segmented videos are also still a challenge.

6.3 AU intensity

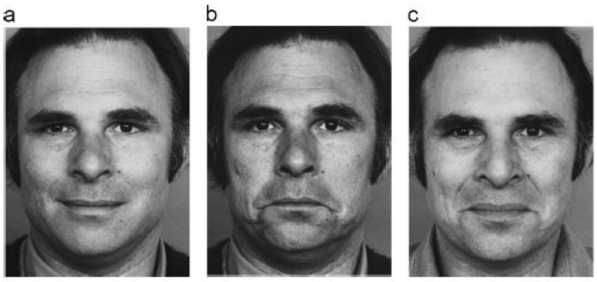

Figure 7: Nonadditive effect in AU combination, (a) AU12 occurs alone. (b) AU15 occurs alone. (c) AU12 and AU15 appear together [30].

AU intensity scoring is done on a five-point ordinal scale, A-B-C-D-E, where A refers to a trace of the action and E to maximum evidence. Measuring the intensity of AUs in a single frame of video is not easy due to the variety, ambiguity and dynamic nature of facial actions. It becomes more difficult for spontaneous facial expressions. When AUs occur in a combination, they may be non-additive. As described before, this means the appearance of an AU in a combination is different from its stand-alone appearance. Fig.7 shows a case of a non-additive AU. To the left (a), the AU12 appears alone, the lip corners are pulled up toward the cheekbone; however, if AU 15 is also becoming active (c), then the lip corners are somewhat angled down due to the presence of AU15. This effect produced from non-additive AUs increases the difficulty of recognizing AUs individually. This is only one example of the problems which are detected when working with AU intensity.

The AU intensity estimation is nowadays posed as a regression problem, categorized in binary classification based, multi-class, regression-based and Ordinal regression. The first three categories are also described in [15], a work which presents the difficulty in estimating the smile intensity. More informations about each class of this regression problem are briefly described in [1].

7 Challenges and opportunities

A lot of work is done in machine analysis of AUs but still some environmental conditions bring big challenges in this field. On the one hand the non frontal head poses, co-occurring AUs and speech, varying illumination conditions and low intensity make these challenges not trivial. On the other hand the small amount of data is another factor which stops the progression. The following text shortly explains some of these challenges and describes the suggestions given and solutions tried in order to solve these problems.

Non frontal head poses: These occur frequently in naturalistic settings. Three possible solutions for this challenge are suggested. The first one gives focus to 3D databases, which may ease this problem, but rendering examples of AU activations at multiple poses still remain not trivial. The second solution suggests the use of head-posed-normalised images for learning. The difficulty in this case is the fact, that the leaning algorithms should be able to cope with partially corrupted data resulting from self-occlusions. The third suggestion is related to the fact that the AUs cause only local appearance changes. This means that a partial occlusion of the face can be problematic. In the paper of Lin et al. [16] attention is given to this problem. A possible solution would be to rely on the semantics of the AUs so that the occluded ones can be inferred and only the visible ones are taken.

AU co-occurrences: The challenge related to AU co-occurrences becomes difficult with the presence of non-additive AUs. One suggestion was made in [18], where these AU combinations are treated in new independent classes. But this method has been shown to be impractical, given the number of such non-additive AU combinations. Another problem appears in the form of a reduction of performance of the algorithms, if these combinations are treated as independent classes [2]. A way to solve the AU co-occurrences problem would be by means of modeling the "semantics" of facial behavior like temporal co-occurrences of AUs. The challenge becomes more trivial knowing that some AU cannot co-occur together. Despite that the problem is still open and needs future researches.

Low intensity AUs: Focus should be given to the situations where the subject is intentionally controlling his facial behavior. For these cases, more scenarios similar to deceit detection would be advantageous. Research in this area is described in the paper [28].

Limitations in databases: An overall list with all available databases related with facial expression recognition is given in the paper from Martinez et al [1]. Hence, all these databases suffer from various limitations, the most important being the lack of realistic illumination conditions and naturalistic head movements. This is why creating publicly available datasets without limitations would be very important and helpful in this field. These problems do not lead to a continued evolution of the field.

8 Conclusion

This paper provides an overview of the different techniques used in order to form a face recognition system. It reaches from face detection to the analysis of AU with the usage of frame or segment based approaches. It shows different works done until now in separate categories of this system. Various features are presented and described, in order to see all the possibilities before starting to build a facial recognition system. In the following some conclusions for each step are presented.

Pre-processing: Starting from face detection, Viola and Jones Algorithms has the best performance and is the most used one. When it comes to facial landmarking, a tracking algorithm is desired, as it can offer much more stable detections. Regression-based methods are nowadays the most robust ones.

Feature Extraction: It is generally advised to use both appearance and geometric features. Simple strategies like feature-level fusion or even decision-level fusion perform well in practice. For their combination, often the Multiple Kernel Learning framework is used. If the best feature is the appearance or the geometric one, cannot be found. Spatio-temporal appearance features provide a consistent and significant advantage, and they can be relatively efficient too. Further use of Deep Learning, in particular CNNs, is an obvious current research focus. How to best combine different features in feature extraction is still an open question.

AU classification, intensity and detection: The best and most effective way to use correlations in practice is to use graph representations. Temporal correlations are easy to obtain and provide important performance improvements. Due to severe label imbalance, it is a good idea to pre-train the frame-based mode chosen and then use a graphical model which takes the output confidence as the input to the graph [1].

Overall, although a major progress in machine recognition of AUs has been made over the years, still there are a lot of challenges open. Many problems are still waiting to be researched. A fully automated system would open potential for applications in security, games, industry and other fields.

References

[1] B.Martinez, M.F.Valstar, B.Jiang and M.Pantic. *Automatic Analysis of Facial Action: A Survey*,2014

[2] B.Jiang, M.F.Valstar, B.Martinez, and M.Pantic. *Dynamic appearance descriptor approach to facial actions temporal modeling.* Trans. Systems, Man and Cybernetics, Part B, pages:161-174, 2014

[3] C.Darwin. *The expression of emotions in man and animal.* John Murray, 1872

[4] C. Liu, J. Yuen, and A. Torralba. *SIFT flow: Dense correspondence across scenes and its applications.*IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 5, pages: 978–994,2011.

[5] C. A. Corneanu, M. Oliu, Jeffrey F. Cohn, and S. Escalera. *Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-related Applications.* IEEE Trans. Pattern Anal. Mach. Intell.

[6] C.Geng and X.Jiang. *Face automatic face recognition framework based on local and global features.* Machine Vision and Application, 2013

[7] Y.Chang and Chang and M.Turk *Manifold of Facial Expression* Image and Vision Computing, Vol.24, No.6, pages:605-614

[8] C.Geng and X.Jiang. *Face recognition based on the multi-scale local image structures.*. Pattern Recognit, 2011

[9] D.Lowe. *Object recognition from local scale-invariant features.* ICCV 2, pages:1150–1157,1999

[10] D.Lowe. *Distinctive image features from scale-invariant keypoints.* IJCV 60, pages:91–110, 2004

[11] E.Sariyanidi, H.Gunes and A.Cavallaro. *Automatic analysis of facial affect: A survey of registration, representation, and recognition.* Trans.Pattern Anal. and Machine Intel, 2015

[12] G.B. Duchene. *Mecanisme de la physionomie humaine, ou analyse electrophysiologique de ses differents modes de l'expression.* Archives generales de medecine, pages:29-47, 1862

[13] G. Tzimiropoulos, V. Argyriou, S. Zafeiriou and T. Stathaki. *Robust FFT-based scale-invariant image registration with image gradients.* IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 10,pages:. 1899–1906, 2010.

[14] Gokturk, S.B, Bouguet, J.Y., Tomasi, C. and Girod,B. *Model-based face tracking for view independent facial expression recognition*Proc. IEEE Int'l Conf. Face and Gesture Recognition, pages:111-116

[15] J. M. Girard, J. F. Cohn, and F. D. la Torre. *Estimating smile intensity: A better way,* Pattern Recognition Letters, vol. 66, 2015

[16] J.-C. Lin, C.-H. Wu, and W.-L. Wei. *Facial action unit prediction under partial occlusion based on error weighted cross-correlation model*, Int'l Conf. Acoust., Speech and Signal Processing, pages: 3482-3486, 2000

[17] M.Yang, D.J.Kriegman and N.Ahuja. *Detecting Faces in Images: A Survey*,2002

[18] M.H.Mahoor, M.Zhou, K.L.Veon, M.Mavadati, and J.F.Cohn *Facial action unit recognition with space representation*,Proc.IEEE Int'l Conf.Automatic Face and Gesture Recognition, pages:336-342, 2011

[19] O.Celiktutan, S.Ulukaya and B.Sankur. *A comparative study of face landmarking techniques*,2013

[20] P.Ekman. *Darwin, deception, and facial expression.* Annuals of the New York Academy of Sciences, 1000: pages:205-221, 2003

[21] P.Ekman, W.V.Friesen. *Facial Action Coding System: A technique for the measurement of facial movement.* Consulting Psychologist Press, 1978

[22] P.Ekman, W.V.Friesen. *Measuring Facial Movement*, 1976

[23] P.Ekman and L.E.Rosenberg. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System.* Oxford University Press, 2005

[24] P.Ekman and K.Scherer. *Handbook of Methods in Nonverbal Behavior Research.* Cambridge Univ. Press; Cambridge,1982.

[25] P.Viola and M.Jones. *Robust real-time object detection.* In Int'l Journal of Computer Vision, 2004

[26] T.F.Cootes, G.J.Edwards and C.J.Taylor. *Active appearance models.* IEEE Trans.Pattern Anal. and Machine Intel, 2001

[27] T.F.Cootes, G.J.Edwards and C.J.Taylor. *Face recognition using active appearance models.* European Conference on Computer Vision, vol. LNCS-Series pages:1406-1607, 1998

[28] TT.Pfister, X.Li, G.Zhao, and M.Pietikainen. *Recognising spontaneous facial micro-expressions.* Int'l Conf. on Computer Vision, pages:14449-1456, 2011

[29] Y.Tian, T. Kanade and J.F. Cohn. *Recognizing Action Units for Facial Expression Analysis*,2001

[30] Y.Li, S.M.Mavadati, M.H.Mahoor, Y.Zhao, and Q.Ji. *Measuring the intensity of spontaneous facial action units with dynamic Bayesian network.* Pattern Recognition, 2015

[31] Y.Tong, J.Chen, and Q.Ji. *A unified probabilistic framework for spontaneous facial action modeling and understanding.* Trans.Patterns Anal. And Machine Intel, 2010

[32] Y.Tong, W.Liao, and Q.Ji. *Facial action unit recognition by exploiting their dynamic and semantic relationships.* Trans Anal.Machine Intel, 2007