A Survey on Facial Feature Extraction to Recognize Facial Expressions

Ms. Priyanka A. Jalan

M.Tech., Dept. of Computer Science & Engg. Bapurao Deshmukh College of Engineering Sevagram, Maharashtra, India

Prof. M. S. Nimbarte

Dept. of Computer Science and Engineering Bapurao Deshmukh College of Engineering Sevagram, Maharashtra, India

Abstract— **The tracking and recognition of facial activities from images or videos have attracted great attention in computer vision field. The facial feature tracking and expression recognition represent the facial activities in three levels from local to global, and they are interdependent problems. However, most current methods only track or recognize the facial activities in one or two levels, and track them separately, either ignoring their interactions or limiting the interaction to one way. The universally accepted categories of emotion are: Sadness, Anger, Happiness, Fear, Disgust and Surprise. In previous days much work has been carried out for facial feature extraction and expression recognition using various techniques for improving the accuracy rate however the aim of proposed work is to achieve high level accuracy reducing the computational time. Instead of the work done yet there is needed to achieve huge level accuracy by reducing the computational time.**

I. INTRODUCTION

The community of computer vision has attracted the attention of facial emotion recognition over the last decade A vast amount of work has been done and is in progress to make life easy for the disabled (e.g. blind, dumb) and aged people by the help of improving all aspects of interaction between computers and human beings. In the area of HCI, there is practical usage based emphasis for automating recognition of a particular facial expression out of a pre-defined list. The recognition of face expression is an important way in interaction between human and computer. Mathematical model of facial expression has been done by extracting the features from face. Many researchers have been proposed to analyze facial expression.

The facial feature tracking and expression recognition represent the facial activities in three levels from local to global, and they are interdependent problems.

For example, facial feature tracking can be used in the feature extraction stage in expression recognition, and expression recognition results can provide a prior distribution for facial feature points. However, most current methods only track or recognize the facial activities in one or two levels, and track them separately, either ignoring their interactions or limiting the interaction to one way.

Human beings possess ability to communicate through facial emotions in every day social interactions. As per application in the HCI field, the universally accepted categories of emotion are: Sadness, Anger, Happiness, Fear, Disgust and Surprise.

However, the most comprehensive system for synthesizing facial expression, developed by Ekman and Friesen, is based on what they called as action units. Facial feature points encode critical information about face shape and face shape deformation. Accurate location and tracking of facial feature points are important in the applications such as animation, computer graphics, etc.

The feature points positions are updated (or projected) simultaneously, which indicates that the interactions within feature points are interdependent. Generally, human faces have a quite sophisticated structure, and a simple parallel mechanism may not be adequate to describe the interactions among facial feature points. The three states of mouth, i.e., widely open, open and closed. However, the discrete states still cannot describe the details of each facial component movement, i.e., only three discrete states are not sufficient to describe all mouth movements

A relatively small subset of six basic expression categories that is found to be recognizable across cultures, for use by FER systems. This set is human universal facial expressions of emotion. This set is described by the emotions: Happy, Sad, Surprised, Anger, Disgust and Fear are considered. There are two broad methods which work by either classifying using a conventional classifier, or act by a fuzzy based possibility driven weighted classification.

II. FACIAL FEATURE EXTRACTION TECHNIQUES

A. Dynamic Bayesian Network(DBN)

In contrast to the mainstream approaches, a probabilistic model based on the Dynamic Bayesian Network (DBN) has been built to capture the facial interactions at different levels. Hence, the flow of information is two-way, not only bottom-up, but also top-down. In general not only the facial feature tracking can contribute to the expression/AUs recognition, but the expression/AU recognition also helps to improve the facial feature tracking performance. All three levels of facial activities are recovered simultaneously through a probabilistic inference by systematically combining the measurements from multiple sources at different levels of abstraction. The facial feature point measurements has been tracked through an active shape model (ASM) based approach [1], which first searches each point locally and then constrains the feature points based on the ASM model, so that the feature points can only deform in specific ways found in the training data. All the 26 facial feature point

positions are manually labeled in each training image.

B. Crop faces & Interpolation

The face is cropped to isolate such regions. The cropping process was done with respect to the nose region and cropping points were chosen along the nose tip vertical and horizontal axes [2]. This cropping process could include regions of the background, as shown in Figure 2(a), with the top left and top right corners representing intersection regions between face and background. The Curvelet transform considers the intersection region as significant regions due to the high variance at edges (curves). To reduce the Curvelet sensitivity to these regions were interpolated face regions over extracted background regions.
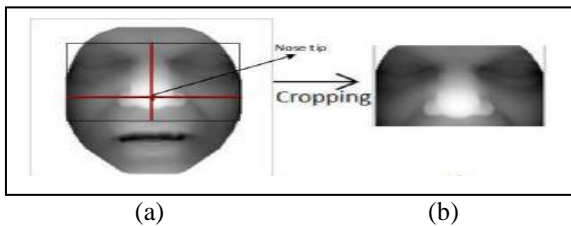

(a)   (b)

Figure 1. a) Face before cropping. b) Face after cropping originally shown in [2]

After the completion of cropping process, each cropped face was decomposed into four levels of scales using the Curvelet transform. Scale 1 captures the lowest frequency components with high variance while Scale 4 captures the highest frequency components with low variance. However, they do not strongly capture geometric features, due to the absence of angle decomposition. Scale 2 and scale 3 have 16 and 32 sub-bands of Curvelet coefficients, respectively. Thus, the Curvelet transform was applied to cropped faces before extracting the features (coefficients) of each wanted region separately.
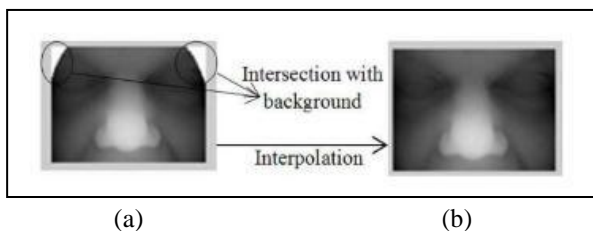

(a)   (b)

Figure 2. a) Cropped face before interpolation. b) Cropped face after interpolation originally or actually shown in [2].

C. Video-Patches

During the offline training phase, the system extracts video patches from different locations of the training videos and separately learns class representatives for each of the six classes [4]. During the online testing phase, a similarity for all extracted video-patches of the query video from class representatives is obtained and a voting based strategy was used to decide about the class of the query face.

An expression video was generally modeled to contain segments such as neutral followed by onset, apex and an offset. Thus modeling the complete video sequence as a whole could result in performance degradation. Therefore,

there is need to extract local video-patches of different lengths from numerous locations of a video.

D. Discriminative Features

Facial feature extraction involves extracting features of facial images; the resulting feature vectors can then be used to project the facial image from the higher dimensional image space into a lower dimensional feature space while preserving the discriminative features [5]. Discriminative features separate the facial images of one class from facial images of other classes in the lower dimensional feature space. Better separation among classes in the lower dimensional space leads to higher classification rate.

E. Region of Interest (ROI)

The ROI strategy gives a better trade-off in eliminating inter AU correlations while modeling AUs and minimizes the constraints on the accuracy of facial feature localization. A feature descriptor that is efficient in representing the facial action units (AUs) while considering only the informative region of interest (ROI) on the face [6]. An in-depth analysis of the state-of-the-art FER techniques under realistic conditions led us to present an ROI strategy which eliminates inter AU correlations in AU modeling.

F. Hybrid Fourier Features

The hybrid Fourier features was extracted from three different Fourier domains in different frequency bandwidths by using a frequency band model selection, and further by adding PCLDA the robustness of the system gets improved. In face recognition, it is not possible to process with the entire extracted features, hence the dimension of the feature vectors has to be reduced [7]. The Fourier features extracted from three different types of domains are concatenated real and imaginary components domain, Fourier spectrum domain, and the phase angle domain.

G. Pseudo Zernike Moments (PZM)

Pseudo Zernike Moments (PZM) is one of the best descriptor that are robust to noise and rotation. Generally PZM was employed to represent faces partitioned into patches. The order parameter specific to each segment, to extract the features of that segment of the input image [12]. Then, the difference between the values of the feature vectors of that segment of the image and the values of the feature vectors of the corresponding segment in the general image is determined to obtain the weight of each segment.

H. Multilayer Histogram of Oriented Gradients (MLHOG)

A new hybrid-based feature called multilayer histogram of oriented gradients (MLHOGs) which adopts the active shape model (ASM) for region extraction pre-processing, and shape characteristics for smiling face detection. The MLHOGs improve the original idea of the histogram of oriented gradients (HOGs) [13] by employing the concept of multilayer's obtained from the ASM and spatial kernels. For classification, linear SVMs are adopted to model the MLHOGs with low computational cost and high accuracy.
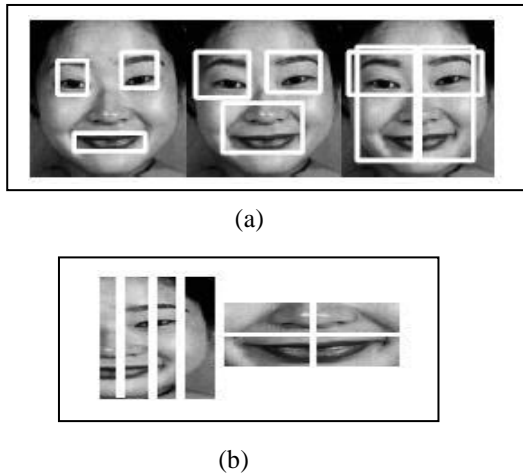
(a)

(b)

Figure 3. a)Examples of predefined regions: from local to global regions. b)Two spatial kernels: pyramid and vertical originally shown in [13]

Fig. 3(a) shows local to global regions for MLHOG features. The local regions focus on the shapes of facial features like the mouth angle and the eye size. The edges caused by the muscles around facial features such as crow's feet and smile folds are involved in the medium regions. In the global regions, the relationship among facial features is the main point.

I. Scale Invariant Feature Transform (SIFT)

SIFT applies the scale-space Difference of Gaussian (DOG) to detect key points in 2D images. The SIFT operator works on each PFI separately. PFIs highlight local shape and texture characteristics of smooth facial images, many more SIFT-based key points can be detected for matching step than those in the original range images [14]. The upper row shows the image is the original facial expression image along with the PFIs in the first two directions. The bottom rows display the PFIs in the remaining six directions.
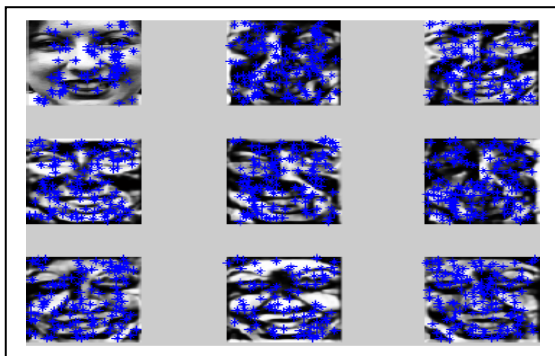

Figure 4. SIFT-based key point detection originally shown in [14].

For example on the Cohn and Kanade data-base, it shows that the average number of key point detected from original facial image is only 91; while using the proposed PFIs for key point detection, the average numbers rise up to 150 for the first direction. Fig. 4 illustrates this phenomenon.

III. CLASSIFICATION TECHNIQUES

A. Principle Component Analysis (PCA)

PCA is one of the most widely used methods in image recognition and compression. PCA aims to obtain a set of mutually orthogonal bases that describe the global information of the data points in terms of variance. PCA has been successfully applied to discover the subspace of the face space, which is termed as eigenfaces.

Scatter maximization is the main drawback in PCA which is not only due to the between-class scatter that is useful for classification, but also to the within- class scatter. Maximization of within-class scatter includes unwanted information to the classification process [2]. Hence, PCA might be optimal in terms of dimensionality reduction; however, it may not be optimal in terms of discriminating images of one class from images of other classes.

B. Grassmannian Manifold

A Grassmannian manifold is a space of all d-dimensional linear subspaces. The existing clustering techniques on the Grassmannian manifold need to compute the distance as well as the mean of the points on the manifold [4], for every iteration of the clustering algorithm (e.g. K-means). Methods of computations of the mean and distance on the Grassmannian manifold can be broadly categorized as intrinsic and extrinsic.

C. Local Fisher Discriminant Analysis

The technique of facial expression recognition in the encrypted domain based on LFDA. Basically, LFDA divides image samples in each class into multiple local classes in the higher dimensional image space [5]. It then projects images belonging to a local class closer to each other while keeping projected images of other local classes apart. However, FLDA and LFDA project samples of each class separately in the feature space as both of them are supervised feature extractors.

Both PCA and FLDA smear the samples of both classes together while LFDA preserves the local structure by projecting them into separable region(s) in the feature space.

D. AU & Spin Support

Effective representation of AUs plays a crucial role on the overall system accuracy and its ability to handle practical challenges like facial point tracking errors. A feature descriptor that provides discriminative representation and rotation (alignment) invariance for each AU. Each AU is described by a set of muscle movements of facial features like brows, eyes, lips, etc., and results in wrinkles, furrows, change in facial feature shapes, etc., at relative angles and positions [6]. It provides a method to incorporate both appearance changes and relative positions of the facial muscle regions.
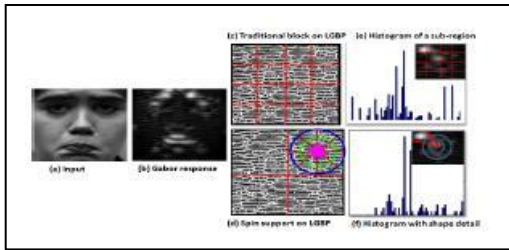
Figure 5. Spin and Rectangle Support Features originally shown in [6]

E. PCLDA

The most traditional subspace analysis methods which are also known as linear methods and its various algorithms are Principal Component Analysis that is PCA [7] and Linear Discriminant Analysis that is LDA. PCA and LDA are the two classic tools widely used in the appearance-based approaches for data reduction and feature extraction. LDA-based algorithms outperform PCA-based ones in low-dimensional representation of the objects. The high dimensional pattern recognition tasks many LDA-based algorithms suffer from the problem called "Small Sample Size problem" (SSS), where the number of available samples is smaller than the dimensionality of the samples. The traditional solution to the SSS problem is to utilize PCA concepts in conjunction with LDA (PCA+LDA) called PCLDA.

F. Active Shape Model (ASM)

The active shape model (ASM) was adopted to locate the discriminative facial features. Moreover, the ASM is more flexible than a conventional statistical model. ASM was employed to simply model the facial image [13]. The ASM can automatically and iteratively locate facial components in the assigned region by using the knowledge learned from trained data. The ASM is trained with the standard annotated face dataset. Although building the ASM requires additional computational costs, using ASMs rather than entire faces can assist the MLHOG module in extracting discriminative and regional characteristics. Besides, there is another benefit that the ASM can easily normalize face without using the eyes finder or other facial feature detectors.

G. Perceived Facial Images (PFIs)

Perceived Facial Images (PFIs), applied to facial images for 2D face recognition. Perceived Facial

Images (PFIs), applied to the problem of face recognition under varying facial expression [14]. PFIs simulate the response of complex neurons to gradient information within a certain neighborhood and possess properties of being highly distinctive as well as robust to affine illumination and geometric transformations.

IV. DATASETS

A. Extended Cohn-Kanade(CK+)

The Extended Cohn Kanade (CK+) dataset [26] is used. The CK+ dataset consists of 593 sequences from 123 posers aged between 18 and 50. The resolution of each sequence is either 640 x 480 or 640 x 490 with 8- bit gray-scale or 24-bit color values. Each sequence has between 10 and 60 frames. The sequences begin with a neutral frame and end in a peak frame [1]. All the sequences are FACS (Facial Action Coding System) coded by human experts who assign labels to the sequences after studying the FACS information in the peak frame of the sequence.

B. SCface

The SCface database contains static images of human faces. Images were captured in uncontrolled indoor environment using five video surveillance cameras of various quality and resolution. The database consists of 4160 face images (in visible and infrared spectrum) of 130 subjects.

An example of normalization in cases of correct/incorrect location is shown in Fig. 6, where the first row shows faces from SCfaces sets Frontal (first image) and R1 (second and third image). As for the first two images, the results by STASM face point location are highlighted [3]. These are correct for the first image but incorrect for the second image. The third image shows the same face as the second face image, with points located according to the position information in the attached file. The second row shows, under each image, the corresponding normalization.
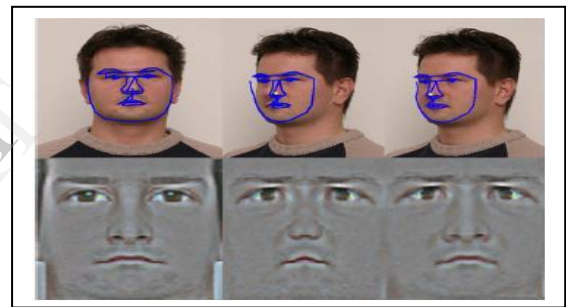

Figure 6. Example faces from SCface normalized by face originally shown in [3]

C. BU4DFE Database

This database comprises 101 subjects (58 females and 43 males) with an age range of 18-45 years, belonging to various ethnical and racial groups including: Asian (28), black(8), Latino(3) and white (6). Under the supervision of a psychologist, each subject is asked to perform 6 different expressions [4]. Each facial expression was captured to produce a 4 seconds video sequence of temporally varying 2D texture and 3D shapes at the rate of 25 frames per second.

D. Japanese Female Facial Expression (JAFFE)

The JAFFE database contains facial images of 10 Japanese females, where each has two to four samples for each expression [14]. In total, there are 213 gray scale facial expression images in this database, each of pixel resolution 256 x 256.

E. Yale B

One is taken in a controlled studio setting while the other is captured in uncontrolled illumination conditions such as hallways, atria, or outdoors [7]. Yale B has provided high resolution images together with four ground

truth locations of the four fiducial points, namely two eyes, nose, and mouth points, to give more chances to improve the recognition performance.

The input image is taken from the Yale B database called test image is shown in Fig.7 (a) and this image has to be recognized in database image called as train image.

(a) Test Image

(b) Gradient   (c) Smoothing

Figure 7. Output of the pre-processed blocks originally shown in [7]

In gradient operation the entire texture information of the test image extracted is shown in Fig.7 (b). In smoothing the illumination sensitivity is reduced by using low pass filter and the smoothed image is shown in Fig.7 (c).

Table 1. Comparison of Different Feature Extraction and Classification Techniques

Feature Extraction	Classification	Datasets	Recognition Rate	Difficulties
Active Shape Model (ASM)	AU Recognition and six basic expression	CK+ & MMI	Average recognition rate is 87.43%	The rigid head movement
Crop faces & Interpolation	PCA	FRGC v2	Non-neutral 93.74% & Neutral 97.+43%	PCA not reduce such dimension
Face Analysis for Commercial Entities (FACE)	FACE has access multiple gallery	Celebrity DB, Labeled Faces in the Wild, Scface. FERET	-	-
Video- patches	Grassmannian manifold	BU4DF	90.97%	Cause the degradation in the overall system
Discriminative Features	LFDA	JAFFE & MUG	JAFFE 94.37% MUG 95.24%	-
Region of Interest (ROI)	AU & Spin-support	CK+, ISL, FACS, JAFFE, MultiPie, MindReading	Whole 57.60% ROI 77.40%	-
Hybrid Fourier Features	PCLDA	Yale B		How to match two faces of same person
Pseudo Zernike Moments (PZM)	Entropy of general images	JAFFE, FG-NET, Radboud	JAFFE 93.12%, FG-NET 93.21%, Radboud 94.51%	-
Fiducial Point,Nearest Class Center (NCC)	Discriminative Isomap,Elastic Body Spline (EBS)	RML emotion,MindReading DVD	D-Isomap 88.2%	-
MLHOG	ASM	JAFFE 7 expression With 213 images	91%	-
Scale Invariant Feature Transform (SIFT)	Perceived Facial Images(PFI)	CK, JAFFE, FEEDTUM (FG-NET)	CK 99.77% JAFFE 96.51% FEEDTUM 98.20%	Expression variation degrade the performance of a face recognition System
Multivariate statistical approaches	Tensor based multivariate statistical approach	Imperial College 3D face & SUNY Binghampton	-	Landmark-dependent may not create realistic facial expression.

V. CONCLUSION

After the investigation of different techniques for facial feature extraction for facial expression recognition, it has been concluded that though different techniques like crop faces and interpolation, Region of Interest, Pseudo Zernike Moments, FACE, discriminative features, video patches etc. used for facial feature extraction for expression recognition provide different level of accuracy but still there is a need to achieve high level of accuracy rate by reducing the computation time.

REFERENCES

[1] Y. Li, S. Wang, et. al. "Simultaneous Facial Feature Tracking & Facial Expression Recognition," IEEE Trans. On Image Proc., Vol. 22, No. 7, pp. 2559-2573, July 2013.

[2] S. Elaiwat, A. El-Sallam, et. al. "3D Face dentification Using Curvelet Transform," IEEE Conf. 2013.

[3] M.D. Marsico, M. Nappi, et. al. "Robust Face Recognition for Uncontrolled Pose and Illumination Changes," IEEE Trans. On Sys.,Man Cybern, Sys., Vol 43, No.1, pp. 149-162, Jan 2013.

[4] M. Hayad, A. El-Sallam, et. al. "Clustering of video- patches on Grassmannian manifold for facial expression recognition from 3D videos," IEEE Conf, pp. 83-88, 2013.

[5] Y. Rahulamatha, W. Phan, et. al. "Facial Expression Recognition in the Encrypted Domain Based on Local Fisher Discriminant Analysis," IEEE Trans. On Affective Comp., Vol 4, No.1, pp. 83-92, Jan-March 2013.

[6] S. Velusamy and B. Anand, "Improved Feature Representation for Robust Facial Action Unit Detection," IEEE CCNC-Work-in-Progress, pp. 681-684, 2013.

[7] Vinothkumar B, Kunar. P, et. al. "A Novel Preprocessing Method & PCLDA Algorithm for Face Recognition Under Difficult Lighting Condition," IEEE Conf. 2013.

[8] Y. Tie, L. Guan, et. al. "A Deformable 3D Facial Expression Model for Dynamic Human Emotional State Recognition," IEEE Trans. On circuit & Sys. For video tech., Vol 23, No.1, pp. 142-157, Jan 2013.

[9] J. B. Kim, Y. Kwang, et. al."Real-Time Realistic 3D Facial Expression Cloning for Smart TV," IEEE International Conf. on Consumer Electronics(ICCE), pp. 240-241, 2013.

[10] X. Zhang, L. Yin, et. al. "A High-Resolution Spontaneous 3D Dynamics Facial Expression Database," IEEE Conf., 2013.

[11] M. Suk, M. Mariappan, et. al. "FaceFetch: A User Emotion Driven Multimedia Content Recommendation System Based on Facial Expression Recognition," IEEE International Symposium on Multimedia, pp. 84-87, 2012.

[12] H. R. Kanan, M. Ahmady, et. al. "Recognition of Facial Expression Using Locally Weighted and Adjusted Order Pseudo Zernike Moments(PZM)," International Conf. on Pattern Recognition(ICPR 2012), pp. 3419-3422, Nov 2012.

[13] H. C. Tsai, W. K. Fan, et. al. "A Real-Time Awareness System for Happiness Expression Based on the Multilayer Histogram of Oriented Gradients," IEEE Conf. 2012.

[14] H. Boughrara, L. Chen, et. al. "Face Recognition under Varying Facial Expression Based on Perceived Facial Images & Local Feature Matching," IEEE International Conf. on Information Tech. & e- Services, 2012.

[15] R. Luo, P. Lin, et. al. "Dynamic Face Recognition System in Recognizing Facial Expression for Service Robotics," IEEE/ASME International Conf. on Advanced Intelligent Mechatronics, pp. 879-884, July 11-14, 2012.

[16] J. L. Minoi, D. F. Jupit, et. al. "Facial Expression Reconstruction of 3D Faces based on real Human Data," IEEE Conf. on Cybern. Com., pp. 185-189, 2012.

[17] K. T. Song, S. C. Chein, et. al. "Facial Expression Recognition Based on Mixture of Basic Expression & Intensities," IEEE Conf. on Sys., Man & Cybern., pp. 3123-3128, Oct 14-17, 2012.

[18] Marc Mehu, Bihan Jiang and Maja Pantic, "Meta- Analysis of the First Facial Expression Recognition Challenge," IEEE Trans. On Sys., Man & Cybern.-Part B: Cybern, Vol 42, No. 4, pp. 966-979, August 2012.

[19] S. Yang, B. Bhanu, et. al. "Understanding Discrete Facial Expression in video using in Emotion Avatar Image," IEEE Trans. On Sys., Man & Cybern., Vol 42, No. 4, pp. 980-992, August 2012.

[20] R. A. Khan, Alexandre Meyer, Hubert Konik and Saida Bouakaz, "Human Vision Inspired Framework For Facial Expressions Recognition," IEEE Conf., pp. 2593-2596, 2012.

[21] S.-W. Lee, D.-J. Kim, K.-H. Park, and Z. Bien, "Gabor wavelet neural network-based facial expression recognition," presented at the World Multi-Conf. Syst. Cybern. Inf., Orlando, FL, 2004.

[22] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn-Kande dataset (CK+): A complete facial expression dataset for action unit and emotion-specified expression," in Proc. 3rd IEEE Int. Conf. Comput. Vis. Pattern Recognit., pp. 94–101, Jun. 2010 .

[23] T. Kanade, J. Cohn, and Y. L. Tian, "Comprehensive database for facial expression analysis," in Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recognit., pp. 46–53, Mar. 2000.

[24] X. Zhao, S. Zhang, et. al. "Facial Expression Recognition Based on Local Binary Patterns and Kernel Discriminant Isomap," Sensors, vol. 11, no. 10, pp. 9573-9588, 2011

[25] J. Wang and L. Yin and X. Wei and Y. Sun. Facial expression recognition based on primitive surface feature distribution. in the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2006), New York, 2006.