© 2018 IJSRSET | Volume 4 | Issue 2 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
National Conference on Advanced Research Trends in Information and
g Technologies (NCARTICT-2018), Department of IT, L. D. College of Engineering,
Ahmedabad, Gujarat, India In association with
International Journal of Scientific Research in Science, Engineering and Technology

A Comprehensive study of Geometric and Appearance based Facial Expression Recognition Methods

Nazeerah A. Sheth[*1], Mahesh M. Goyani[2]
[*1]Department of Computer Engineering, Government Engineering College, Modasa, Gujarat, India
[2]Department of Computer Engineering, Government Engineering College, Modasa, Gujarat, India

ABSTRACT

It is a well-known fact that facial expressions are one of the key reflectors of the emotional state of a person and the research on the same has been spanning for a long time. Being an essential requirement in Human Computer Interaction as well as other applications such as automobile safety, mental health detection, animations, etc. recognizing facial expressions with precision has become vital. This paper presents a survey on various important and effective techniques present in literature along with their variations used recently. Prominent techniques of each step and a detailed discussion on feature extraction methods have been provided along with a detailed comparison of few recent approaches.

Keywords: Facial Expressions, Survey, Emotion Recognition

I. INTRODUCTION

Facial Expressions are one of the most well-known types of non-verbal communication [1]. A single expression can perhaps convey more information than thousands of words combined as it reflects the emotional state of a person which a human brain can efficiently recognize and interpret. Automatic Facial Expression Recognition (AFER) intends to transfer a certain, if not full level of that ability to computers. With the advancements and increasing use of Human-Computer Interaction (HCI) in many applications, the need for effective AFER is all the more necessary.

Apart from socially sensitive Human-Computer Interaction [2], AFER can be utilized in many applications such as detection of mental disorders [3], safety against road rage [4], security [5], animations and video games [6], automate applications, etc. It can also be used to include emotion-related information in automatic image captioning systems [7].

Earlier, the facial expression related study was confined to psychology, medical, artistic or acting fields [8], but with the availability of high computing resources, new and smart technologies, and increasing research in image processing and machine learning, it has garnered a significant amount of interest from computer scientists as well [9]. The study of AFER can be traced back to Darwin, who in his popular book "The Expression of the Emotions in Man and Animals" [10] established general principles of expression and grouped them into various categories, for example, anger and fear combined in one category. He also classified the facial deformations that occur within each category. In 1971, Ekman and Friesen classified emotions into six main categories, namely angry, sad, happy, fear, disgust and surprise [11] as shown in Fig. 1 [12]. To this date, these six emotions are considered to be universal across different races, cultures, gender and ethnicities [13]. With the availability of high computational power and advances in fields of Computer Vision and Robotics, a lot of work has been done on AFER since the 1990's [9].

AFER can be looked at from various perspectives like real-time vs non-real time, geometric vs appearance based, message vs sign judgement, global vs local, static vs dynamic, etc. Some of the significant work done on AFER seems to be biased towards non-real time systems as there are various issues such occlusions, pose variation, low-intensity input, identification of baseline frame, etc. that holds back the performance levels in real-time [14].

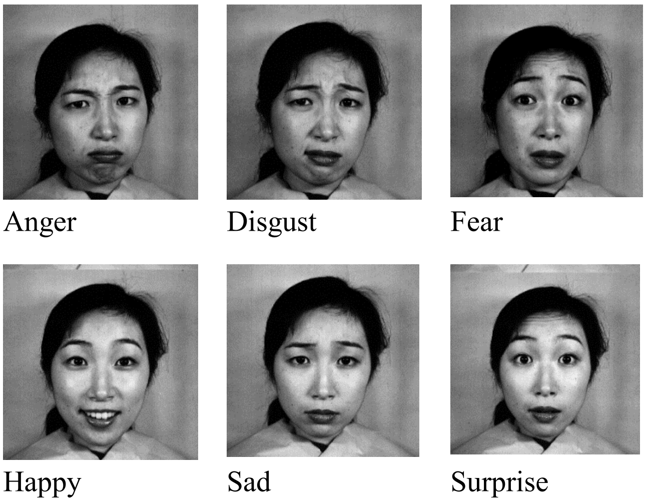

Figure 1. Basic six prototypical expressions

Geometry based approaches consider the position, curvature and deformation information of facial components to extract facial features whereas the appearance based approaches filter the intensity values. Also, both these approaches can be combined in order to combine their advantages. Both these methods will further be discussed in detail in section 3. Features extracted could either be holistic, i.e. generated from the whole face or part based, i.e. generated from specific region of interests like eyes, eyebrows, lips etc. [15].

Irrespective of the perspective followed, a basic AFER system comprises of mainly the stages as shown in Figure 2 [16]. Pre-processing includes enhancing the image in order to facilitate the further process and reduce any such possible conditions that might affect the recognition performance. It also includes Face detection and tracking which involves detecting the face region from the frame and tracking it across subsequent frames. The next stage deals with extracting the expression related features from the face followed by reducing them in order to filter out redundant features and decrease the complexity. The last step includes feeding the features to a classifier which then narrows down to one of the expression classes. The detailed survey on facial expression recognition can be found in [8], [17], [18].

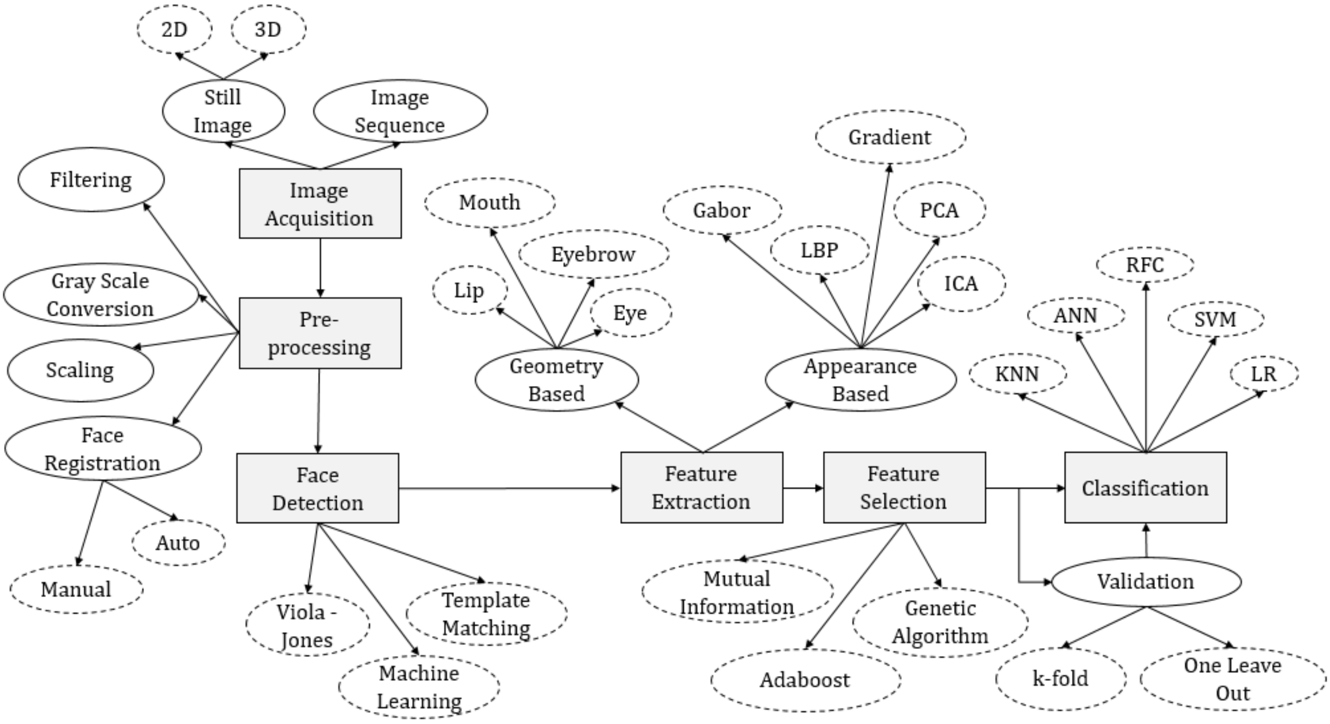

Figure 2. Basic AFER system

The rest of the paper is structured as follows. Section 2 (State Of The Art Methods) contains an in-depth review of various feature extraction techniques including both geometric and appearance based methods followed by Section 3 (Comparison and Analysis) which provides a detailed comparison of each step of various approaches.

Section 4 (**Conclusion and Future Scope**) concludes with various challenges and future scope of the system.

II. STATE OF THE ART METHODS

Feature extraction is probably the most important step in the FER process as it involves identifying and extracting discriminating features that contribute towards various facial expressions. Good features maximize between class dissimilarity and minimize within-class differences and are easily extracted with lower dimensions [19]. These features can be mostly found in the regions around eyes, mouth, nose, and face edges [20]. Methods that extract features relative to the whole face coordinates are known as global methods, while those that extract features relative to inner facial features or regions of interest are known as local methods [21], [15].

Most methods, such as [22], [23], [24], [25], [26] use global feature extraction methods so that a complete set of features are used for better classification purpose. However, approaches like [27], [20], [28], [29] used local techniques so that any redundant or unusable information is not taken into consideration. Another reason for choosing local methods is that they are not affected by face geometry, ageing, varying pose and face rotation [8]. Also, they are computationally less expensive as fewer number of features have to be processed. However, care has to be taken in choosing regions of interest so that vital discriminative features are not lost. Both these techniques can also be combined together as in [30], [31], [32] to balance the tradeoffs. Both local and global methods can generally be classified into geometric and appearance based techniques, an example of which is shown in Figure 3.

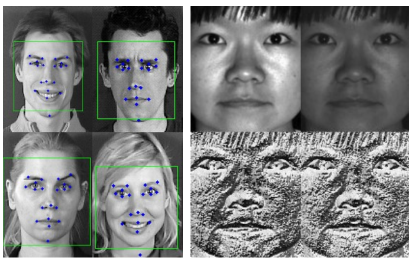

Figure 3. Geometry (left) and Appearance (right) based features

A. Geometric Based Feature Extraction

The Facial Action Coding System (FACS) [33] developed by Ekman and Friesen determined Action Units (AUs) based on landmark facial points whose movements were traced to define facial expressions. Action Unit is a measurement of smallest possible change in facial muscle. There are 44 Action Units defined consisting of various parts of a face, some of which are shown in Figure 4. Based on this approach, a lot of geometric methods have been used for the recognition of facial expressions.

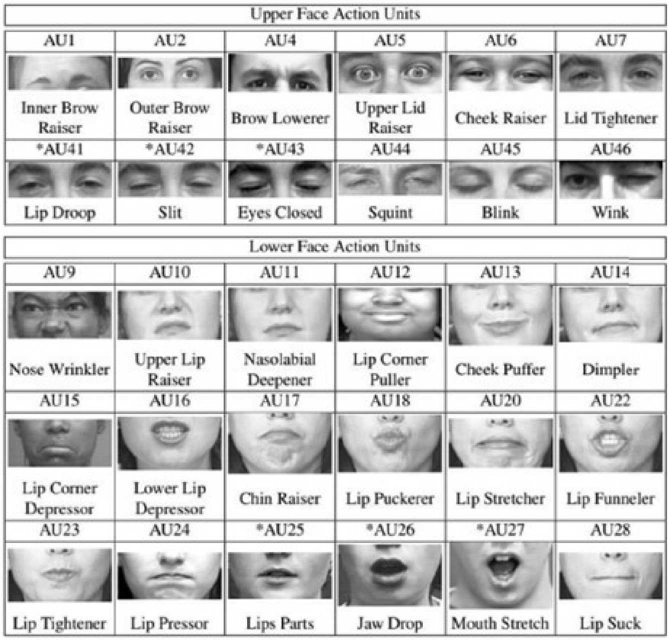

Figure 4. Action Units of upper and lower face parts

Geometric methods do not focus on face texture or intricate settings, rather they take feature indications from the geometry, deformation and tracking of fiducial points [16]. Usually, the methods require a reference to a baseline frame. Features are then extracted by tracking the landmarks through subsequent frames. For example, [34] uses triangles formed by 3 out of 68 landmarks, inspired by polygon mesh representation. The changes in angles and areas of these triangles are evaluated between consecutive frames to extract features. Ghimire et al. [35] used face graph normalization to bring all graphs to normal shape before extracting points, lines and triangles. These are determined by initializing and tracking facial points through a suitable technique.

Since few years, many researches based on 3D facial geometry have been proposed owing to the robustness of 3D image to scale, pose and illumination variations [36]. Zeng et al. [37] proposed a unique surface conformal representation in the form of Conformal Factor Image (CFI) and Mean Curvature Image (MCI). A conformal map is computed and normalized which preserves the angles of facial geometry along with 3 main landmark points which are then used to extract features. Lemaire et al. [38] use 3D depth maps to generate Differential Mean Curvature Maps (DMCP) which improves multi-scale facial expression surface topology uniqueness. The DMCP are then normalized and Histogram of Gradients (HOG) [39] algorithm is used to generate features. Berretti et al. [36] identifies a set of facial key points and computes Scale Invariant Feature Transform (SIFT) [40] feature descriptors from depth images of the face around key points.

B. Appearance Based Feature Extraction

Geometric methods face many shortcomings such as dependency on face geometry and pose, failure to detect landmarks in case of occlusions, no tolerance against face representation errors, etc. [16] Owing to the various pitfalls of geometric methods, appearance based feature extraction methods have gained much popularity in terms of higher accuracy and lower error conditions [16]. Appearance based features take cue from the intensity levels of the face image to determine features contributing to facial expression. These features are extracted by generating a suitable filter which is the convolved around the face [21], [15], [16].

Among the pool of various appearance features, the ones with binarized local texture have shown promising results and effectiveness [21], [19]. In 1996, Ojala et al. [41] formulated the Local Binary Patterns (LBP) which is an efficient and easy non-parametric method to describe mild and intricate texture information and summarize the local structure of facial components. Earlier it was used for face description by Ahonen et al. [42] and texture description by Ojala et al. [43]. Later, many LBP based facial expression recognition systems have been [44], [45], [46], [47], [48]. Huang et al. [19] provides a comprehensive study on various LBP techniques on facial image analysis along with their variations. LBP is robust to monotonic illumination change and misalignment [21] and is computationally simple [49]. However, since the original LBP operates on a 3×3 neighbourhood, it may miss out vital structural information [50]. Therefore Improved LBP was proposed to include neighbourhood of all sizes to capture complete information [43].

Local Directional Pattern (LDP) is another technique proposed by Jabid et al. [51], to address the non–robustness problem of LBP towards random noise and non-monotonic illumination. It consists of directional information by comparing edge response values of each pixel in eight directions using Kirsch masks representing the impact of edge in each direction. The feature shows superior performance and can be represented in low-dimensional feature space with high accuracy for even low-resolution images. However, the LDP codes can be problematic in smooth regions as they focus more on edge response values and produces inconsistent patterns in uniform regions [52]. Jun et al. determined a new robust local gradient coding (LGC) which further implied that gradient differences among pixels improved accuracy [53]. Local Directional Pattern Variance (LDPv) was developed by Kabir et al. [54] to include contrast information along with edge response values by introducing variance as an adaptive weight to modify the LDP code.

Local Ternary Patterns [55] with an extra discrimination level and ternary codes were introduced to address limitations of LDP and also to tackle non-uniform noise. Ahmed et al. [56] proposed Gradient Local Ternary patterns (GLTP) which combines the advantages of LTP and LBP by encoding more robust gradient magnitude values in a three-level scheme to achieve consistent texture patterns in random noise and varying illumination. An improved version of GLTP with a better gradient operator and dimensionality reduction technique was used by Holder et al. [20] which is shown to give better results.

Recently, Ryu et al. [26] defined Local Directional Ternary Pattern (LDTP) which efficiently encodes edge directional information of emotion-related features in edge regions and neglects the less significant smooth regions based on magnitude encoded within the ternary pattern. It is shown to be a highly discriminable and robust pattern, but the recognition rates drop significantly with varying camera angles and ethnic variations [57]. Arshid et al. [31] proposed a Multi Stage Binary Pattern (MSBP) which aims to tackle real-world recognition issues such as varying features, intricate settings and complex backgrounds by retaining local texture variations along with gradient changes along edges such as eyes, wrinkles etc.. In MSBP, a multi-stage binary code is generated for each comparison against neighborhood pixel.

Recently, Goyani and Patel presented two local appearance based FERS techniques: Local Mean Binary Operator (LMBP) [58] and Multi-Level Haar wavelet (MLH) based FERS [59]. MLBP computes the 256 bin histogram of the LBP code of the mean centred 3 × 3 patch, whereas MLH extracts approximation coefficients of the image at multiple scales using Haar wavelets. Haar are inherently good at noise suppression where as LMBP achieves the robustness to noise and non-monotonic change in illumination.

Gabor filter is a linear parametric filter which analyses whether there are any specific frequency content in the image in specific directions in a localized region around the region of analysis [60]. Gabor features suffer from identity bias [16] and hence used for identity recognition in many places [61], [62]. Lyons et al. [61] first used Gabor wavelets to generate facial expression features. It has proved to be an effective method and has since been used by many different approaches like in [63], [64], [65], [66], [67], for facial expression recognition. Local Phase Quantization (LPQ) was a technique used by [68] for blur insensitive texture classification which computes Short Term Fourier Transform (STFT) on local image window. It is then quantized using a scalar binary quantizer for calculating the phase information. Though it is computationally expensive than LBP, Dhall et al. [69] showed that it gives better recognition rates for facial expressions recognition. Histogram of Oriented Gradients (HOG) [70] is a descriptor that counts the occurrences of localized gradient orientation in images and is used for object recognition since it is efficient in encoding shape information. PHOG is an extension to HOG in which gradients in each grid are joined at pyramid level which is shown to give better performances. PHOG have been used for static facial expression analysis in [71] and [72]. To address the problems of high dimensionality and low variance between expression classes, another common technique employed for feature extraction is Fisherfaces Linear Discriminant Analysis (FLDA) [73] in which the discriminating features of an image are preserved with reduced dimensions. It intends to reduce intra class variance and increase inter class variance but fails in case of multimodal classes [73]. To overcome this, Local Fisher Discriminant Analysis (LFDA) was introduced [74], [75]. In this, local between class variance is increased while minimizing the local within-class variance. But it fails to determine essential mixed structure when face image space is highly nonlinear [25]. In order to solve the variance problem along with reduced dimensions, several more methods are present like Kernel Discriminant Analysis (KDA) [76], General Discriminant Analysis (GDA) [77] and Linear Discriminant Analysis (LDA) [76]. LDA is more widely used but is less flexible when comes to complex datasets [25]. Siddiqi et al. [25] used an enhanced version of LDA called Stepwise Linear Discriminant Analysis (SWLDA) which is computationally less expensive with higher predictive ability and gives better results. Another line of approach seen in literature is to cascade two or more effective methods one after the other. In this, one method is applied over the response of the other. One such pattern, LGBP (Local Gabor Binary Pattern) was proposed by Zhang et al. [78] in which Gabor features capture orientations and scales while LBP focuses subtler texture details. LGBP is robust to variations in lighting and expressions. Among others, LGBP was also used in [79] where expressions were classified into dominant and complementary emotions like happily surprised to cover a wider range of emotions instead of basic prototypes. In a recent approach, Sun et al. proposed a novel method in [23] in which instead of applying LBP over the features described by Gabor, both Gabor and LBP features are extracted from the input image and then those features are fused together using feature fusion to form the final feature vector. The approach has yielded superior performance since features described using different perspectives are used together. Similar method was proposed in [80] where Global features extracted by PCA (discussed under feature reduction) are fused with local features extracted from the mouth area using LBP. PCA is an effective method as it reduces dimensionality while selecting distinct features, but the features extracted by it are subject to environment changes [80]. This is combatted by fusing in LBP which focuses on texture details and hence the results show improved robustness. Since LBP is sensitive to random noise and non-monotonic illumination, this method was improved by Luo et al. in [32] where instead of LBP, LDP is used to extract local features from the eyes and mouth region. The method shows improved performance than PCA and LBP combined, since LDP has good stability against random noise. Rajesh et al. experimented various fusion of different methods in [81] such as fusion of LBP with LGC, HOG with LDP, and HOG with wavelets. The results showed that fusion of HOG with wavelets outperformed all others and fusion of LDP with HOG improved LDP's performance. LMBP operator presented by Goyani and Patel [58] is computationally efficient compared to LDP as well it is robust to noise and illumination.

So far, the methods we have discussed work on static facial images with no temporal information. Image sequences can be used to incorporate temporal information which may represent expressions from onset to peak to further offset. Static images usually just include the peak expression image. Image sequences help to incorporate expression at different levels and also makes it easier to differentiate different expressions and have shown better descriptions as shown by [82]. The only drawback is that it takes a toll on the computation as more number of images have to be processed.

Few of the methods discussed earlier have been extended to include spatiotemporal information. LBP was extended to LBP-TOP [82] to include multiple frames. To avoid high computations, feature vectors from only three orthogonal planes namely, XY, XT and YT are concatenated to describe facial expressions [21]. The XY plane gives the static information while temporal information is provided by the other two [83]. Similarly, LPQ and LGBP are also extended to LPQ-TOP [83] and LGBP-TOP [84]. But as indicated in [27], these kind of methods lack: 1). temporal correspondences among different phases of an expression, and 2). semantic representation, since these low-level representations do not convey any semantic meaning of individual parts of an expression. For example, if the indicated expression is happy, then the low-level representation do not convey a smiling component. To overcome both these shortcomings, [27] have proposed a mid-level representation through expressionlets. Expressionlets aim to bridge the gap between low level representations and high level semantics by modelling each video clip as a Spatio Temporal Manifold (STM) composed of low level features. A Uniform Manifold Model (UMM) is built based on all STMs and each expression is then fitted as an instance of the UMM. Owing to the novel solution, this method has shown significant improvement over other video based recognition methods, but still lacks accuracy for challenging real-time datasets. Spatio Temporal Texture Map (STTM) [22] was developed to capture subtle spatial and temporal variations of expressions with low computational cost. It does so by extracting information from 3D Harris Corner function and representing the features in the form of histograms. The model is better suited for real-time expressions considering the lower computational cost. Compared to expressionlets it gives much better performance recognizing the disgust emotion, but the accuracy for others falls slightly behind. Another dynamic expression recognition approach was developed recently by Kamarol et al. [85] which couples intensity estimation along with expression recognition with low level computation. The model is based on the fact that each person expresses emotions at different intensities and thus a system that adapts to individual person would be useful in applications like pain detection from videos.

III. COMPARISON AND ANALYSIS

The previous section evaluated and analyzed various methods for feature extraction starting LBP and its variations developed over the years. These methods have shown great performance in representing mild texture details and have managed to overcome various challenges pertaining to effective FER. Parametric methods that use frequency and phase information have also shown to be effective. The effectiveness of methods that combine two or more techniques in different ways indicates that right combination of techniques can significantly improve recognition rate. FER systems based on image sequences are used to include time-based information and different levels of expressions and have shown better performance.

In this section we present a comparison of various FER approaches on various parameters and techniques used for each step. Table 1 provides the FER systems with geometry based feature extraction whereas Table 2 shows the systems with appearance based feature extraction.

Table 1. Comparision Of Geometry Based Facial Expression Recognition Methods

Ref.	PP	FE	FS	S/D	L/G	Classifier	NOE	Val	Dataset	P/S	Accuracy
[24]	-	ASM	-	S	G	Double SVM, sample selected SV	6	-	JAFFE	P	98.25
[34]	-	Angles, areas by landmark points	Exhaustive analysis of 68 landmarks	D S	G	CRF c-kNN (citation Nearest neighbour)	7	LOO	CK+	P	86.7 82.2
[35]	Face graph normalization, V&J, EBGM, KLT– point detection & track	Point, line, triangles	FS AdaBoost	D	G	SVM	6	10F CD	CK+ MMI MUG	P S P	97.80 77.22 95.50
[38]	Align model to frontal pose, cropping	DMCM HOG	DMCM Norm	D	G	SVM	7	10F	BU-3DFE	P	76.6
[36]	-	SIFT	mRMR	S	G	SVM	7	10F	BU-3DFE	P	78.43
[85]	Procrustes analysis	AAM	-	D	G	HMM	6	SI 10F LOO	CK+ BUFED	P P	82.4 62.86

Table 2. Comparision Of Appearance Based Facial Expression Recognition Methods

Ref	PP	FE	DR	FS	S/D	G/L	Classifier	NOE	Val	Dataset	P/S	Accuracy
[46]	Re-projected from 3D model to 5 images of different angles, V&J	LBP and variations	-	-	S	L G	SVM	6	10F	BU-3DFE MPIE	P	67.96 73.26
[86]	Regression Trees for FD	LBP NCM	ES	-	S	L	SVM	6 7	5F	CK+	P	94.83 97.25
[23]	Rotation, BG crop, Histogram equilization, AdaBoost	Gabor + LBP	PCA	LDA	S	G	SVM	6 7	10F	CK+	P	97.42(6) 95.45(7)
[27]	Normalization	Expressionlet	PCA	DL	D	L	Multi class linear SVM	6 7	PI 10F	CK+, MMI, CASIA, AFEW	P P P S	94.19 75.12 74.59 31.73
[22]	BG crop, V&J	STTM (3D harris corner)	-	-	D	G	Multi class SVM	6	2F	CK+, CASME II, AFEW	P S S	97.70 98.61 90.68
[25]	-	SWLDA	-	-	D	G	Hidden CRF	6	10F nF LOO	CK+ JAFFE B+ MMI	P P P P	96.37

[32]	Normalization, Adaboost detection	PCA + LDP	-	-	S	G L	SVM	7	LOO	-	P	91.61
[31]	Normalization, V & J	MSBP	Divide MSBP into 2 parts	-	S	G L	BF tree Simple- logistic kNN Bagging NB	7	10F	SFEW	S	96 (G) 60 (L)
[87]	Illumination enhancement, Face location, Normalization BG crop, V & J	LGBP	-	-	D	L	SVM	30+	Votin g	CK+ JAFFE	P	73.88
[26]	Normalization	LDTP	-	-	S	L	SVM	7	NF LOO	CK+ JAFFE MMI CMU-PIE GEMEP-F BU-3DFE	P	94.2 94.8 99.8 89.5 98.1 88.1
[20]	Crop face & components, Landmark locations and V&J	Improved GLTP	PCA	-	S	L	SVM	6 7	10F LOO PI	CK+ JAFFE	P P	95.17 85.05
[67]	Normalization, Integral projection of horizontal and vertical direction to detect eye lips	Gabor	-	FSDD	S	L	Local weighted AdaBoost	2	-	JAFFE CK	P P	96.17 95.10
[66]	Normalization, V & J	Gabor Template	-	M- FS, TM	S	G	Norm based SVM	7	10F	CK JAFFE	P P	95.1- 90.8 80.3 78.4
[83]	Procrustes Transform, Automatic point detector	LPQ- TOP	-	GB	D	G	SVM MM	AUs	nF CD	MMI UNBCSA L GEMEP- FERA CK SEMAINE	P S S P P S	-
[84]	Normalization, Local Evidence Aggregated Regression (LEAR) detect points	LGBP- TOP	-	F1 based FS	D	G	SVM	AUs	10F SI CV	MMI CK	P P	-

PP- Pre Processing, FE- Feature Extraction, DR- Dimensionality Reduction, FS- Feature Selection, S-Static, D-Dynamic, L-Local, G-Global, NOE-Number of Expressions, P-Posed, S-Spontaneous, Val-Validation, BG-Background, nF- n-Fold Cross Validation, CD- Cross-Database, PI- Person Independent, LOO- Leave One Out

IV. CONCLUSION

This paper presented a review of the various state of the art techniques used for facial expression recognition. The survey starts with a quick description about the importance, applications and some background information about facial expression recognition systems, followed by a take on various perspectives an AFER system can be looked at. Various steps of a basic AFER system are then discussed including the assessment of numerous methods and techniques corresponding to each step. An in-depth analysis of various feature extraction methods has been provided including their motives and shortcomings.

The study has shown that despite a lot of research and achievements in the literature, there is still scope for improvement considering the dynamic nature of facial expressions and various challenges associated with it. For an AFER system to be effectively implemented, we need the system to be robust and accurate. This precision can be reached if all the challenges that reduce the efficiency are tackled. As it is observed that different techniques address different problems, it makes sense that a certain type of combination of some of them would achieve the desired goal. It is also deduced that instead of just the prototypical expressions, the range of expressions can be widened to cover more detailed and subtle emotions.

V. REFERENCES

[1]. A. Mehrabian, "Communication without words," Psychology Today, vol. 2. pp. 53-55, 1968.

[2]. F. Cid, J. A. Prado, P. Bustos, and P. Nunez, "A Real Time and Robust Facial Expression Recognition and Imitation approach for Affective Human-Robot Interaction Using Gabor filtering," IEEE International Conference on Intelligent Robots and Systems, pp. 2188-2193, 2013.

[3]. W. Gaebel and W. Wolwer, "Facial expression and emotional face recognition in schizophrenia and depression," European Archives of Psychiatry and Clinical Neuroscience, vol. 242, no. 1, pp. 46-52, 1992.

[4]. F. Nasoz, K. Alvarez, C. L. Lisetti, and N. Finkelstein, "Emotion recognition from physiological signals using wireless sensors for presence technologies," Cognition, Technology & Work, vol. 6, no. 1, pp. 4-14, 2004.

[5]. A. Butalia, M. Ingle, and P. Kulkarni, "Facial Expression Recognition for Security," International Journal of Modern Engineering Research, vol. 2, no. 4, pp. 1449-1453, 2012.

[6]. M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan, "Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction," IEEE Conference on Computer Vision and Pattern Recognition Workshop, vol. 5, pp. 53-53, 2003.

[7]. R. Cowie et al., "Emotion recognition in human-computer interaction," Signal Processing Magazine, IEEE, vol. 18, no. 1, pp. 32-80, 2001.

[8]. B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," Pattern Recognition, vol. 36, no. 1, pp. 259-275, 2003.

[9]. V. Bettadapura, "Face Expression Recognition and Analysis : The State of the Art," arXiv preprint arXiv:1203.6722, pp. 1-27, 2012.

[10]. Charles Darwin, The expression of the emotions in man and animals. Oxford University Press, USA, 1998.

[11]. P. Ekman, "Universal and Cultural differences in facial expressions of emotions," Nebraska Symposium on Motivation, vol. 19. pp. 207-282, 1971.

[12]. M. J. Lyons, "Automatic classification of single facial images," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 12, pp. 1357-1362, 1999.

[13]. P. Ekman, "Strong Evidence for universals in facial expressions," Psychol. Bull, vol. 115, no. 2. pp. 268-287, 1994.

[14]. S. Deshmukh, M. Patwardhan, and A. Mahajan, "Survey on real-time facial expression recognition techniques," IET Biometrics, vol. 5, no. 3, pp. 243-251, 2016.

[15]. C. A. Corneanu, M. O. Simón, J. F. Cohn, and S. E. Guerrero, "Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-Related Applications," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 8, pp. 1548-1568, 2016.

[16]. E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition,"

IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 6, pp. 1-22, 2015.
[17]. M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions : the state of the art," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1424-1445, 2000.
[18]. M. Goyani, "A Comprehensive Study on Texture Feature Based Facial Expression Recognition," 2017.
[19]. D. Huang, C. Shan, M. Ardebilian, and L. Chen, "Facial Image Analysis Based on Local Binary Patterns : A Survey," IEEE Transactions on Image Processing, 2011.
[20]. R. P. Holder and J. R. Tapamo, "Improved gradient local ternary patterns for facial expression recognition," EURASIP Journal on Image and Video Processing, vol. 2017, no. 1, pp. 42-57, 2017.
[21]. B. Martinez, M. F. Valstar, B. Jiang, and M. Pantic, "Automatic Analysis of Facial Actions : A Survey," IEEE Transactions on Affective Computing, 2017.
[22]. S. K. A. Kamarol, M. H. Jaward, J. Parkkinen, and R. Parthiban, "Spatiotemporal feature extraction for facial expression recognition," IET Image Processing, vol. 10, no. 7, pp. 534-541, 2016.
[23]. Y. Sun and J. Yu, "Facial Expression Recognition by Fusing Gabor and Local Binary Pattern Features," International Conference on Multimedia modeling. Springer, pp. 209-220, 2017.
[24]. T. Yu and X. Gu, "Facial Expression Recognition Using Double-Stage Sample-Selected SVM," International Conference on Intelligent Computing, Springer, pp. 304-315, 2017.
[25]. M. H. Siddiqi, R. Ali, A. M. Khan, Y. T. Park, and S. Lee, "Human Facial Expression Recognition Using Stepwise Linear Discriminant Analysis and Hidden Conditional Random Fields," IEEE Transactions on Image Processing, vol. 24, no. 4, pp. 1386-1398, 2015.
[26]. B. Ryu, A. Ramirez Rivera, J. Kim, and O. Chae, "Local Directional Ternary Pattern for Facial Expression Recognition," IEEE Transactions on Image Processing, vol. 26, no. 12, pp. 6006-6018, 2017.
[27]. M. Liu, S. Shan, R. Wang, and X. Chen, "Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2014, pp. 1749-1756.
[28]. D. Ghimire, S. Jeong, J. Lee, and S. H. Park, "Facial expression recognition based on local region specific features and support vector machines," Multimedia Tools and Applications, vol. 76, no. 6, pp. 1-19, 2016.
[29]. A. Uçar, Y. Demir, and C. Güzeliş, "A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering," Neural Computing and Applications, vol. 27, no. 1, pp. 131-142, 2016.
[30]. A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," IEEE Winter Conference on Applications of Computer Vision, pp. 1-10, 2016.
[31]. S. Arshid, A. Hussain, A. Munir, A. Nawaz, and S. Aziz, "Multi-stage binary patterns for facial expression recognition in real world," Cluster Computing, pp. 1-9, 2017.
[32]. Y. Luo, T. Zhang, and Y. Zhang, "A novel fusion method of PCA and LDP for facial expression feature extraction," International Journal for Light and Electron Optics, vol. 127, no. 2, pp. 718-721, 2016.
[33]. P. Ekman and W. Friesen, Manual for the Facial Action Coding System. Consulting Psychologist Press, Palo Alto, 1977.
[34]. D. Acevedo, P. Negri, M. E. Buemi, F. G. Fernandez, and M. Mejail, "A Simple Geometric-Based Descriptor for Facial Expression Recognition," 12th IEEE International Conference on Automatic Face & Gesture Recognition, pp. 802-808, 2017.
[35]. D. Ghimire, J. Lee, Z. N. Li, and S. Jeong, "Recognition of facial expressions based on salient geometric features and support vector machines," Multimedia Tools and Applications, vol. 76, no. 6, pp. 7921-7946, 2016.
[36]. S. Berretti, B. Ben Amor, M. Daoudi, and A. Del Bimbo, "3D facial expression recognition using SIFT descriptors of automatically detected keypoints," Visual Computer, vol. 27, no. 11, pp. 1021-1036, 2011.

[37]. W. Zeng, H. Li, L. Chen, J.-M. Morvan, and X. D. Gu, "An automatic 3D expression recognition framework based on sparse representation of conformal images," 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, pp. 1-8, 2013.
[38]. P. Lemaire, M. Ardabilian, L. Chen, and M. Daoudi, "Fully automatic 3D facial expression recognition using differential mean curvature maps and histograms of oriented gradients," 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, pp. 1-7, 2013.
[39]. Dahmane, Mohamed, and J. Meunier, "Emotion recognition using dynamic gridbased HoG features," in IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, 2011, pp. 884-888.
[40]. Lowe and G. David, "Distinctive image features from scale-invariant key points," International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.
[41]. T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," Pattern Recognition, vol. 29, no. 1, pp. 51-59, 1996.
[42]. T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: application to face recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pp. 2037-2041, 2006.
[43]. T. Ojala, M. Pietikäinen, and T. Mäenpää, "A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification," Advances in Pattern Recognition, pp. 399-408, 2001.
[44]. C. Shan, S. Gong, and P. W. Mcowan, "Facial expression recognition based on Local Binary Patterns : A comprehensive study," Image and Vision Computing, vol. 27, no. 6, pp. 803-816, 2009.
[45]. C. Shan, S. Gong, and P. W. McOwan, "Robust facial expression recognition using local binary patterns," IEEE International Conference on Image Processing, vol. 2, pp. 170-173, 2005.
[46]. S. Moore and R. Bowden, "Local binary patterns for multi-view facial expression recognition," Computer Vision and Image Understanding, vol. 115, no. 4, pp. 541-558, 2011.
[47]. R. Hablani, N. Chaudhari, and S. Tanwani, "Recognition of facial expressions using Local Binary Patterns of important facial parts," International Journal of Image Processing, vol. 7, no. 2, pp. 163-170, 2013.
[48]. X. Zhao and S. Zhang, "Facial expression recognition using local binary patterns and discriminant kernel locally linear embedding," EURASIP Journal on Advances in Signal Processing, vol. 2012, no. 1, p. 20, 2012.
[49]. D. Huang, C. Shan, M. Ardabilian, Y. Wang, and L. Chen, "Local Binary Patterns and Its Application to Facial Image Analysis: A Survey," IEEE Transactions on Systems Man and Cybernetics Part C-Applications and Reviews, vol. 41, no. 6, pp. 765-781, 2011.
[50]. H. Jin, Q. Liu, H. Lu, and X. Tong, "Face detection using improved LBP under Bayesian framework," in International Conference on Image and Graphics, 2004, pp. 306-309.
[51]. T. Jabid, M. H. Kabir, and O. Chae, "Facial expression recognition using Local Directional Pattern (LDP)," 17th IEEE International Conference on Image Processing, pp. 1605-1608, 2010.
[52]. F. Ahmed and M. H. Kabir, "Directional ternary pattern (dtp) for facial expression recognition," in IEEE International Conference on Consumer Electronics, 2012, pp. 265-266.
[53]. B. Jun and D. Kim, "Robust face detection using local gradient patterns and evidence accumulation," Pattern Recognition, vol. 45, no. 9, pp. 3304-3316, 2012.
[54]. H. Kabir, T. Jabid, and O. Chae, "Local Directional Pattern Variance (LDPv):Robust Feature Descriptor for Facial Expression Recognition," The International Arab Journal of Information Technology, vol. 9, no. 4, pp. 382-392, 2012.
[55]. X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in Analysis and Modeling of Faces and Gestures, 2007, pp. 168-182.
[56]. F. Ahmed and E. Hossain, "Automated facial expression recognition using gradient-based ternary texture patterns," Chinese Journal of Engineering, pp. 1-8, 2013.

[57]. R. Ahmed, A. Meyer, H. Konik, and S. Bouakaz, "Framework for reliable , real-time facial expression recognition for low resolution images," Pattern Recognition Letters, vol. 34, no. 10, pp. 1159-1168, 2013.
[58]. M. Goyani and N. Patel, "Robust Facial Expression Recognition using Local Mean Binary Pattern," Electronic Letters on Computer Vision and Image Analysis, vol. 16, no. 1, pp. 54-67, 2017.
[59]. M. Goyani and N. Patel, "Multi-Level Haar Wavelet based Facial Expression Recognition using Logistic Regression," Indian Journal of Science and Technology, vol. 10, no. 9, pp. 1-9, 2017.
[60]. G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 10, pp. 974-989, 1999.
[61]. M. Lyons and S. Akamatsu, "Coding Facial Expressions with Gabor Wavelets," 3rd IEEE Conference on Automatic Face and Gesture Recognition, pp. 200-205, 1998.
[62]. M. Lades et al., "Distortion Invariant Object Recognition in the Dynamic Link Architecture," IEEE Transactions on Computers, vol. 42, pp. 300-311, 1993.
[63]. W. Gu, C. Xiang, Y. V. Venkatesh, D. Huang, and H. Lin, "Facial expression recognition using radial encoding of local Gabor features and classifier synthesis," Pattern Recognition, vol. 45, no. 1, pp. 80-91, 2012.
[64]. F. Chen and K. Kotani, "Facial Expression Recognition by SVM-based Two-stage Classifier on Gabor Features," IAPR Conference on Machine Vision Applications, pp. 453-456, 2007.
[65]. M. J. Lyons, J. Budynek, A. Plante, and S. Akamatsu, "Classifying Facial Attributes using a 2-D GaborWavelet Representation and Discriminant Analysis," in 4th International Conference on Automatic Face and Gesture Recognition, 2000, pp. 202-207.
[66]. L. Zhang, D. Tjondronegoro, and V. Chandran, "Random Gabor based templates for facial expression recognition in images with facial occlusion," Neurocomputing, vol. 145, pp. 451-464, 2014.
[67]. Y. Xing and W. Luo, "Facial Expression Recognition Using Local Gabor Features and Adaboost Classifiers," IEEE Internation al Conference on Progress in Informatics and Computing, pp. 1-5, 2016.
[68]. V. Ojansivu and J. Heikkil, "Blur insensitive texture classification using local phase quantization," Image and Signal Processing, pp. 236-243, 2008.
[69]. A. Dhall, A. Asthana, R. Goecke, and T. Gedeon, "Emotion recognition using PHOG and LPQ features," IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, pp. 878-883, 2011.
[70]. N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," International Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 886-893, 2005.
[71]. Z. Li, J. Imai, and M. Kaneko, "Facial-component-based Bag of Words and PHOG Descriptor for Facial Expression Recognition," in International Conference on Systems, Man, and Cybernetics San, 2009, pp. 1353-1358.
[72]. Y. Bai, L. Guo, L. Jin, and Q. Huang, "A novel feature extraction method using pyramid histogram of orientation gradients for smile recognition," in 16th IEEE International Conference on Image Processing, 2009, pp. 3305-3308.
[73]. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs . Fisherfaces : Recognition Using Class Specific Linear Projection," vol. 19, no. 7, pp. 711-720, 1997.
[74]. M. Sugiyama, "Dimensionality Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis," Journal of Machine Learning Research, vol. 8, pp. 1027-1067, 2007.
[75]. Y. Rahulamathavan, R. C. Phan, J. A. Chambers, and D. J. Parish, "Facial Expression Recognition in the Encrypted Domain Based on Local Fisher Discriminant Analysis," vol. 4, no. 1, pp. 83-92, 2013.
[76]. S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Muller, "Fisher Discriminant Analysis with Kernels," in IEEE Signal Processing Soceity Workshop Neural Network Signal Process IX, 1999, pp. 41-48.
[77]. G. Baudat and F. Anouar, "Generalized Discriminant Analysis using a Kernel Approach,"

Neural Computing, vol. 12, no. 10, pp. 2385-2404, 2000.

[78]. W. Zhang, S. Shan, W. Gao, and X. Chen, "Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition," IEEE 10th International Conference on Computer Vision, vol. 1, pp. 786-791, 2005.

[79]. C. Loob et al., "Dominant and Complementary Multi-Emotional Facial Expression Recognition Using C-Support Vector Classification," IEEE 12th International Conference on Automatic Face & Gesture Recogition, pp. 833-838, 2017.

[80]. Y. Luo, C. Wu, and Y. Zhang, "Facial expression recognition based on fusion feature of PCA and LBP with SVM," International Journal for Light and Electron Optics, vol. 124, no. 17, pp. 2767-2770, 2013.

[81]. J. Kumari, R. Rajesh, and A. Kumar, "Fusion of features for the effective facial expression recognition," International Conference on Communication and Signal Processing, pp. 457-461, 2016.

[82]. G. Y. Zhao and M. Pietikainen, "Dynamic Texture Recognition using Local Binary Pattern with an Application to Facial Expressions," Transactions on Pattern Analysis and Machine Intelligence, vol. 2, no. 6, pp. 915-928, 2007.

[83]. B. Jiang, M. Valstar, B. Martinez, and M. Pantic, "A dynamic appearance descriptor approach to facial actions temporal modeling," IEEE Transactions on Cybernetics, vol. 44, no. 2, pp. 161-174, 2014.

[84]. T. R. Almaev and M. F. Valstar, "Local gabor binary patterns from three orthogonal planes for automatic facial expression recognition," in Humaine Association Conference on Affective Computing and Intelligent Interaction, 2013, pp. 356-361.

[85]. S. K. A. Kamarol, M. H. Jaward, H. Kalviaineninen, J. Parkkinen, and R. Parthiban, "Joint facial expression recognition and intensity estimation based on weighted votes of image sequences," Pattern Recognition Letters, vol. 92, pp. 25-32, 2017.

[86]. D. Ghimire, S. Jeong, J. Lee, and S. H. Park, "Facial expression recognition based on local region specific features and support vector machines," Multimedia Tools and Applications, vol. 76, no. 6, pp. 7803-7821, 2017.

[87]. C. Loob et al., "Dominant and Complementary Multi-Emotional Facial Expression Recognition Using C-Support Vector Classification," IEEE 12th International Conference on Automatic Face & Gesture Recogition, pp. 833-838, 2017.

[88]. H. Naveen Kumar, S. Jagadeesha, and A. K. Jain, "Human Facial Expression Recognition from static images using shape and appearance feature," 2nd International Conference on Applied and Theoretical Computing and Communication Technology, pp. 598-603, 2016.

[89]. F. Bashar, A. Khan, F. Ahmed, and M. H. Kabir, "Robust facial expression recognition based on median ternary pattern (MTP)," International Conference on Electrical Information and Communication Technology, pp. 1-5, 2013.