Importance of Emotion Recognition Using Image Processing – A Literature Survey

Dr. P. Sumathy #1, Ahilan Chandrasekaran *2

[1]*Assistant Professor, Department of Computer Science, Bharathidasan University,*
[1]*Tiruchirappalli, Tamil Nadu, INDIA – 620024*
[2]*Research Scholar, Department of Computer Science, Bharathidasan University,*
[2]*Tiruchirappalli, Tamil Nadu, INDIA – 620024*
[1]ahilanchandrasekaran@gmail.com
[2]sumathy.p@bdu.ac.in

***Abstract*— Emotion recognition based on facial expression is an important field in affective computing. Current emotion recognition systems may suffer from two shortcomings: translation in facial image may deteriorate the recognition performance, and the classifier is not robust. Affective computing is currently one of the most active research topics, furthermore, having increasingly intensive attention. This strong interest is driven by a wide spectrum of promising applications in many areas such as virtual reality, smart surveillance, perceptual interface, etc. Affective computing concerns multidisciplinary knowledge background such as psychology, cognitive, physiology and computer sciences. This paper is emphasized the literature survey on the facial emotion recognition using Image Processing techniques by various researchers in the year 2013-2017. Various methods are discussed in order to examine the state of the art. Finally, some research challenges and future directions are also discussed.**

***Keywords*— Affective Computing, Facial Expression, Image Processing, Emotion Recognition.**

I. Introduction

Affective computing enable computers have human-like capabilities such of observation, interpretation and generation of affect features. It is an important topic for harmonious human-computer interaction by improving the quality of human computer communication and the intelligence of the computer. The research on affect or emotion is from 19th century [1]. Traditionally, "affect" is seldom linked to lifeless machines and is normally studied by psychologists. It is quite new in the recent years that the affect features have captured and processed by the computer. The affective computing builds an "affect model" based on various sensors-captured information and builds a personalized computing system with the capability of perception, interpretation to human's feeling as well as giving intelligent, sensitive and friendly responses. To get the impression of state of art of the research in affective computing, the paper briefly summaries few key technologies in the research during last several years, such as facial expression, emotion recognition using Image Processing techniques. The rest of the paper is organized as follows. The literature survey is presented in the section 2. Section 3 explains the survey findings and the research direction is presented in the section 4. The paper is concluding in the last section.

A. Facial Expression based Emotion Recognition

Facial expressions and movements such as a smile or a nod are used either to fulfil a semantic function to communicate emotions or as conversational cues. Similar as speech processing, the research of facial expression consists of works on coding, recognition as well as generation and have been used for a long time. For instance, Etcoff [2] parameterized the structure of the chief parts of human's face through 37 lines, which enables people to roughly capture the affect status of faces. Ekman [3] built facial action coding system. He has classified human's facial expressions into many action units. This method has described facial expressions with six basic emotions such as joy, anger, surprise, disgust, fear and sadness. Currently, most of the facial features can be found from the definition of MPEG-4. The MPEG-4 allows the user to configure and build systems for many applications by allowing flexibility in the system configurations, by providing various levels of interactivity with audio-visual content [4][5]. In this standard, both mesh model [6] or muscle model are used to create 3-D facial models. To perform the facial expression analysis, most of the facial features have been captured by the optical flow or active appearance model. Lyons [7] have applied the supervised Fisher Linear Discriminant Analysis (FDA) to learn an adequate linear subspace from class specified training samples and the samples projected to this subspace can be best separated. Principal Component Analysis (PCA) [8] and Independent Component Analysis (ICA) [9] have been used for the expression classification. For facial expression recognition, there are many methods, such as, Gabor wavelets [7], neural network [10], Hidden Markov Models (HMM) [5], Point Distribute Model (PDM), optical flow, geometrical tracking method, EGM method and so on. Among them, the Gabor representation has been favored by many researchers due to its performance and in sensitive to the face posture and the lighting background. Many researchers used methods based on images [11][12], visemes [13], FAPs [14], PCs [15][16], 3D coordinates [13], 3D distance measurements [5] [15] or optical flows [17] to generate facial expression purposes.

B. Body Gesture and Movement based Emotion Recognition

Body gesture and movement is defined by the positions of body arthroses and their changes with time. Currently, the work for gesture processing is more focused on the hand tracking. Hand gestures convey various and diverse meanings, to enhance the mood or behaviour as a sign language. Traditionally, there are two methods and they are apparentness [18] and 3D modeling [19]. The apparentness based method makes out the model by analyzing apparent features of hand gestures from 2D images, while the 3D methods do the tracking in real-time 3D environment. Compared to 3D methods, the apparentness methods are less complicated, and easier to be used in real-time computation. However requires more effects to adapt into high noise background and the real time application. Some efforts have been made to adopt mixed modeling methods and describe the features of static hand gesture with multiple features (such as local profile features and overall image matrix features) [20]. It shows the higher and more robust tracking results.

In order to realize body gesture and movement based from image sequences, the key point is how to confirm the positions of body arthroses according to given image information. The existing methods normally have some limitation [19], such as, different dress colours according to different body arthroses, simple moving directions, simple backgrounds, some manual initial markings. With these methods, the profile of the target body is picked up at first, and then virtual framework that is similar to real body framework is taken out through energy function. Arthroses positions are determined based on the virtual framework by using anthroponomy knowledge. The energy function can restrain some background noises, and has low requirement for the preciseness of the fetched body profile. In addition, some people enable the computers to more accurately capture data of face and body's rapid movement by some auxiliary equipment like electromagnetic inductor [21] and optical reflection signs [22]. Recent past, the work is still a difficult subject in computer vision's area, especially in real application. Concerning the capture of body gesture and movement, in addition to further improvement of the capture accuracy and efficiency of parameters, how to obtain more robust and subtle body-language is still an urgent difficult problem for affective computing.

II. Literature Survey

Liu, Mengyi, et al [23] have proposed a method for video-based human emotion recognition. For each video clip, all frames are represented as image set, which can be modeled as a linear subspace to be embedded in Grassmannian manifold. Class-specific One-to-Rest Partial Least Squares (PLS) is learned on video and audio data respectively to distinguish each class from the other confusing ones. Ghimire et. al [24] have presented method for fully automatic facial expression recognition in facial image sequences.

The prototypical expression sequence for each class of facial expression is formed, by taking the median of the landmark tracking results from the training facial expression sequences. Multi-class AdaBoost with dynamic time warping similarity distance between the feature vector of input facial expression and prototypical facial expression is used as a weak classifier to select the subset of discriminative feature vectors. Finally, two methods for facial expression recognition are presented, either by using multi-class AdaBoost with dynamic time warping, or by using Support Vector Machine (SVM) on the boosted feature vectors. Kahou, Samira Ebrahimi, et al [25] have presented the techniques used for the University of Montreal's team submissions 2013 Emotion Recognition in the Wild Challenge. The proposed approach combines multiple deep neural networks for different data modalities, including: a deep convolutional neural network for the analysis of facial expressions within video frame, a deep belief net to capture audio information, a deep autoencoder to model the spatio-temporal information produced by the human actions within the entire scene along with a shallow network architecture focused on extracted features of the mouth of the primary human subject in the scene.

Zavaschi, Thiago HH, et al [26] have presented a novel method for facial expression recognition that employs the combination of two different feature sets in an ensemble approach. A pool of base support vector machine classifiers is created using Gabor filters and Local Binary Patterns. A multi-objective genetic algorithm is used to search the best ensemble as objective functions the minimization of both the error rate and the size of the ensemble. Danisman, Taner, et al [27] have presented an automatic way to discover pixels in a face image that improves the facial expressions recognition results. The proposed method exhaustively searches for the best and worst feature window position from a set of face images among all possible combinations using Multi-Layer Perceptron (MLP). Kalita, Jeemoni, and Karen Das [28] have proposed an Eigenvector based system has been presented to recognize facial expressions from digital facial images. In the approach, firstly the images have been acquired and cropped five significant portions. Wagner, Jennifer B., et al [29] have examined the neural, behavioral and autonomic correlates of emotional face processing in adolescents with ASD and typical development (TD) using eye-tracking and event-related potentials (ERPs) across two different paradigms. Scanning of faces was similar across groups in the first task, but the second task found that face-sensitive ERPs varied with emotional expressions only in TD. Zhang, Li, et al [30] have proposed a work that concentrates on intelligent neural network based facial emotion recognition and Latent Semantic Analysis based topic detection for a humanoid robot. The work has first of all incorporated Facial Action Coding System (FACS) describing physical cues and anatomical knowledge of facial behaviour for the detection of neutral and six basic emotions from real-time posed facial expressions. Feedforward Neural Networks (NN) are used to respectively implement both upper and lower facial Action Units (AU) analysers to recognise six upper and 11 lower facial actions including Inner and Outer Brow Raiser, Lid Tightener, Lip Corner Puller, Upper Lip Raiser, Nose Wrinkler, Mouth Stretch etc. Rivera, et.al [31] proposed a novel local feature descriptor, Local Directional Number Pattern (LDN), for face analysis and expression recognition. LDN encodes the directional information of the face’s textures (i.e., the texture’s structure) in a compact way, producing a more discriminative code than current methods. Sikka, Karan, et al [32] proposed a method to automatically detect emotions in unconstrained settings as part of the 2013 Emotion Recognition in the Wild Challenge. The method combines multiple visual descriptors with paralinguistic audio features for multimodal classification of video clips. Extracted features are combined using Multiple Kernel Learning and the clips are classified using SVM into one of the seven emotion categories: Anger, Disgust, Fear, Happiness, Neutral, Sadness and Surprise. Ahsan, Tanveer, et.al [33] have proposed an approach in pursuit of recognizing facial expression where facial feature is represented by a hybrid of Gabor wavelet transform of an image and local transitional pattern code. Expression images are classified into prototype expressions via support vector machine with different kernels. Sadeghi, et.al [34] have presented an approach, which eliminates geometric variability in emotion expression. Thus, appearance features can be accurately used for facial expression recognition. Therefore, a fixed geometric model is used for geometric normalization of facial images. This model is defined as one of the emotional expressions. In addition, Local Binary Patterns are utilized to represent facial appearance features. Cid, Felipe, et al [35] have presented a real-time system for recognition and imitation of facial expressions in the context of affective Human Robot Interaction. The proposed method achieves a fast and robust facial feature based on consecutively applying filters to the gradient image. Gabor filter is used along with a set of morphological and convolutional filters to reduce the noise and the light dependence of the image acquired by the robot. A set of invariant edge-based features are extracted and used as input to a Dynamic Bayesian Network classifier in order to estimate a human emotion. The output of this classifier updates a geometric robotic head model, which is used as a bridge between the human expressiveness and the robotic head. Young, et.al [36] have presented an approach to determine whether child maltreatment has a long-term impact on emotion processing abilities in adulthood and whether IQ, psychopathology, or psychopathy mediate the relationship between childhood maltreatment and emotion processing in adulthood. Majumder, et.al [37] have presented a novel emotion recognition model using the system identification approach. A comprehensive data driven model using an extended Kohonen self-organizing map (KSOM) has been developed whose input is a 26 dimensional facial geometric feature vector comprising eye, lip and eyebrow feature points. Owusu, Ebenezer, Yongzhao Zhan, and Qi Rong Mao [38] has improved the recognition accuracy and execution time of facial expression recognition system will be used. The face detection component is implemented by adopted Viola–Jones descriptor. The detected face is down-sampled by Bessel transform to reduce the feature extraction space to improve processing time. Gabor feature extraction techniques were employed to extract thousands of facial features which represent various facial deformation patterns. An AdaBoost-based hypothesis is formulated to select a few hundreds of the numerous extracted features to speed up classification. Zhang, Ligang, Dian Tjondronegoro, and Vinod Chandran [39] proposed a robust approach that employs a Monte Carlo algorithm to extract a set of Gabor based part-face templates from gallery images and converts these templates into template match distance features. The resulting feature vectors are robust to occlusion because occluded parts are covered by some but not all of the random templates. Bejani, Mahdi, Davood Gharavian, and Nasrollah Moghaddam Charkari [40] have simulated human perception of emotion through combining emotion-related information using facial expression and speech. Speech emotion recognition system is based on prosody features, mel-frequency cepstral coefficients (a representation of the short-term power spectrum of a sound) and facial expression recognition based on integrated time motion image and quantized image matrix, which can be seen as an extension to temporal templates. Owusu, Ebenezer, Yonzhao Zhan, and Qi Rong Mao [41] have focused on improving the recognition rate and processing time of facial recognition systems. The skin is detected by pixel-based methods to reduce the searching space for Maximum Rejection Classifier (MRC) to detect the face. The detected face is normalized by a Discrete Cosine Transform (DCT) and down-sampled by Bessel transform. Gabor feature extraction techniques were utilized to extract thousands of facial features that represent facial deformation patterns. An AdaBoost-based hypothesis is formulated to select a few hundreds of Gabor features which are potential candidates for expression recognition. Vitale, Jonathan, et al [42] used Simulation Theory and neuroscience findings on Mirror-Neuron Systems as the basis for a novel computational model, as a way to handle affective facial expressions. The model is based on a probabilistic mapping of observations from multiple identities onto a single fixed identity, and then onto a latent space.

Ali, Hasimah, et al [43] have proposed a new method of using empirical mode decomposition (EMD) technique for facial emotion recognition. The EMD algorithm can decompose any nonlinear and non-stationary signal into a number of intrinsic mode functions (IMFs). In this method, the facial signal obtained from successive projection of Radon transform of 2D image is decomposed using EMD into oscillating components called IMFs. The first IMF (IMF1) was extracted and considered as features to recognize the facial emotions. Three dimensionality reduction algorithms: Principal Component Analysis (PCA) + Linear Discriminant Analysis (LDA), PCA + Local Fisher Discriminant Analysis (LFDA), and

Kernel LFDA (KLFDA) were independently applied on EMD-based features for dimensionality reduction. These dimensionality reduced features were fed to the k-Nearest Neighbor (k-NN), Support Vector machine (SVM) and Extreme Learning Machines with Radial Basis Function (ELM-RBF) classifiers for classification of seven universal facial expressions. Rice, Linda Marie, et al [44] have examined the extent to which a computer-based social skills intervention called FaceSay was associated with improvements in affect recognition, mentalizing, and social skills of school-aged children with Autism Spectrum Disorder (ASD). Hossain, M. Shamim, and Ghulam Muhammad [45] have proposed a cloud-assisted speech and face recognition framework for elderly health monitoring, where handheld devices or video cameras collect speech along with face images and deliver to the cloud server for possible analysis and classification. In the framework, a patient's state such as pain, tensed, and so forth is recognized from his or her speech and face images. Ming, Yue [46] presented a robust regional bounding spherical descriptor (RBSR) to facilitate 3D face recognition and emotion analysis. In this framework, the authors first segment a group of regions on each 3D facial point cloud by shape index and spherical bands on the human face. The corresponding facial areas are projected to regional bounding spheres to obtain regional descriptor. Finally, a regional and global regression mapping (RGRM) technique is employed on the weighted regional descriptor for boosting the classification accuracy. Palo, H. K., Mihir Narayana Mohanty, and Mahesh Chandra [47] proposed an attempt to recognise two classes of speech emotions as high arousal like angry and surprise and low arousal like sad and bore. Linear Prediction Coefficients (LPC), Linear Prediction Cepstral Coefficient (LPCC), Mel Frequency Cepstral Coefficient (MFCC) and Perceptual Linear Prediction (PLP) features are used for emotion recognition using Multi-Layer Perception (MLP).Various emotional speech features are extracted from audio channel using above-mentioned features to be used in training and testing. Barros, Pablo, et al [48] have used a hierarchical feature representation to deal with spontaneous emotions, and learns how to integrate multiple modalities for non-verbal emotion recognition, making it suitable to be used in Human – Robot Interaction (HRI) scenario. Zhang, Li, et al [49] have proposed unsupervised automatic facial point detection integrated with regression-based intensity estimation for facial action units (AUs) and emotion clustering to deal with such challenges. The proposed facial point detector is able to detect 54 facial points in images of faces with occlusions, pose variations and scaling differences using Gabor filtering, BRISK (Binary Robust Invariant Scalable Key points), an Iterative Closest Point (ICP) algorithm and fuzzy c-means (FCM) clustering. Neoh, Siew Chin, et al [50] have proposed a facial expression recognition system with a layered encoding cascade optimization model. Since generating an effective facial representation is a vital step to the success of facial emotion recognition, a modified Local Gabor Binary Pattern operator is first employed to derive a refined initial face representation and use two evolutionary algorithms for feature optimization including direct similarity and Pareto-based feature selection, under the layered cascade model. Mlakar, Uroš, and Božidar Potočnik [51] have proposed an efficient automated method for facial expression recognition based on the histogram of oriented gradient (HOG) descriptor. This subject-independent method was designed for recognizing six prototyping emotions. It recognizes emotions by calculating differences on a level of feature descriptors between a neutral expression and a peak expression of an observed person. The parameters for the HOG descriptor were determined by using a genetic algorithm. Support Vector Machines (SVM) was applied during the recognition phase, where one SVM classifier was trained for one emotion. Liu, Mengyi, et al [52] have proposed to construct a deep architecture. AU inspired Deep Networks (AUDN), inspired by the psychological theory that expressions can be decomposed into multiple facial Action Units (AUs). To fully exploit this inspiration but avoid detecting AUs, proposed to automatically learns informative local appearance variation, optimal way to combining local variation and high level representation for final expression recognition. Boughrara, Hayet, et al [53] have presented a constructive training algorithm for Multi-Layer Perceptron (MLP) applied to facial expression recognition applications. The developed algorithm is composed of a single hidden-layer using a given number of neurons and a small number of training patterns. The MLP constructive training algorithm seeks to find synthesis parameters as the number of patterns corresponding for subsets of each class to be presented initially in the training step. The initial number of hidden neurons, the number of iterations during the training step as well as the MSE predefined value. Kahou, Samira Ebrahimi, et al [54] have presented an approach for learning several specialist models using deep learning techniques, where each focuses on one modality. Among these are, a convolutional neural network, focusing on capturing visual information in detected faces, a deep belief net focusing on the representation of the audio stream, a K-Means based "bag-of-mouths" model, which extracts visual features around the mouth region and a relational auto encoder, which addresses spatio-temporal aspects of videos. Manos, Leandro Y., et al [55] have discussed the use of patient images and emotional detection to assist patients and elderly people within an in-home healthcare context. In addition, there are few studies that consider patient's emotional state, which is crucial them to be able to recover from a disease. Alhussein, Musaed [56] have proposed an emotion recognition system from face for e-Healthcare system. Weber Local Descriptors (WLD) are utilized as feature in this approach, a static facial image is subdivided into many blocks. A multiscale WLD is applied to each of the blocks to obtain a WLD histogram from the image. The significant bins of the histogram are found by using Fisher discrimination ratio. These bin values represent the descriptors of the face. The face descriptors are then input to a support vector machine to recognition the emotion.

Kim, Bo-Kyeong, et al [57] have described an approach towards robust Facial Expression Recognition (FER) for the third Emotion Recognition in the Wild challenge. The authors trained multiple deep Convolutional Neural Networks. The authors presented two strategies to obtain diverse decisions from deep CNNs. The network architecture, input normalization, and random weight initialization in training these deep models may be changed. A hierarchical architecture of the committee with exponentially-weighted decision fusion is constructed. Sun, Bo, et al [58] designed two temporal–spatial dense scale-invariant feature transform (SIFT) features and combine multimodal features to recognize expression from image sequences. For the static facial expression recognition based on video frames, we extract dense SIFT and some deep convolutional neural network (CNN) features, including the proposed CNN architecture. Calvo, Manuel G., and Lauri Nummenmaa [59] proposed the visual and emotional factors in expression recognition system. The behavioral, neurophysiological, and computational measures indicate that basic expressions are reliably recognized and discriminated from one another. The effect may be inflated by the use of prototypical expression stimuli and forced-choice responses. The affective content along the dimensions of valence and arousal is extracted early from facial expressions. However, this coarse affective representation contributes minimally to categorical recognition of specific expressions. The physical configuration and visual saliency of facial features contribute significantly to expression recognition, with "emotionless" computational models being able to reproduce some of the basic phenomena demonstrated in human observers. Zhang, Li, et al [60] have proposed a facial expression recognition system with a variant of evolutionary fire- fly algorithm for feature optimization. Modified Local Binary Pattern descriptor is proposed to produce an initial discriminative face representation. A variant of the firefly algorithm is proposed to perform feature optimization. The proposed evolutionary firefly algorithm exploits the spiral search behaviour of attractiveness search actions of fireflies to mitigate premature convergence of the Levy-flight firefly algorithm (LFA) and the moth-flame optimization (MFO) algorithm. Zhang, Xiao, and Mohammad H. Mahoor [61] have presented a method, task-dependent multi-task multiple kernel learning (TD-MTMKL) to jointly detect the absence and presence of multiple AUs. TD-MTMKL attempts to learn an optimal kernel combination from a given set of basis kernels for each involved AU and obtain a finer depiction of AU relations through kernel combination weights. Altamura, Mario, et al [62] have presented an Emotional face recognition is impaired in bipolar disorder, but it is not clear whether this is specific for the illness. Here, the authors have investigated how aging and bipolar disorder influence dynamic emotional face recognition. Twenty older adults, 16 bipolar patients, and 20 control subjects performed a dynamic affective facial recognition task and a subsequent rating task. Kar, Nikunja Bihari, Korra Sathya Babu, and Sanjay Kumar Jena [63] have proposed a new hybrid system for automatic facial expression recognition. The proposed method utilizes histograms of oriented gradients (HOG) descriptor to extract features from expressive facial images. Feature reduction techniques namely principal component analysis (PCA) and linear discriminant analysis (LDA) are applied to obtain the most important discriminant features. Finally, the discriminant features are fed to the back-propagation neural network (BPNN) classifier to determine the underlying emotions from expressive facial image.

Liu, Zhentao, et al [64] depicted the facial expression emotion recognition based human-robot interaction (FEER-HRI) system is proposed, for which a four-layer system framework is designed. The FEERHRI system enables the robots not only to recognize human emotions, but also to generate facial expression for adapting to human emotions. A facial emotion recognition method based on 2D-Gabor, uniform local binary pattern (LBP) operator, and multiclass extreme learning machine (ELM) classifier is presented, which is applied to real-time facial expression recognition for robots. Lopes, André Teixeira, et al [65] have proposed a simple solution for facial expression recognition that uses a combination of Convolutional Neural Network and specific image pre-processing steps. Convolutional Neural Networks achieve better accuracy with big data. Fan, Xijian, and Tardi Tjahjadi [66] have focused on basic emotion recognition and propose a spatiotemporal feature based on local Zernike moment in the spatial domain using motion change frequency. The authors also design a dynamic feature comprising motion history image and entropy. To recognise a facial expression, a weighting strategy based on the latter feature and sub-division of the image frame is applied to the former to enhance the dynamic information of facial expression, and followed by the application of the classical support vector machine. Datta, Samyak, Debashis Sen, and R. Balasubramanian [67] have presented a fast facial emotion classification system that relies on the concatenation of geometric and texture-based features. For classification, the authors proposed to leverage the binary classification capabilities of a support vector machine classifier to a hierarchical graph-based architecture that allows multi-class classification. Wen, Guihua, et al [68] have presented an ensemble of convolutional neural networks method with probability-based fusion for facial expression recognition, where the architecture of CNN was adapted by using the convolutional rectified linear layer as the first layer and multiple hidden maxout layers. It was constructed by randomly varying parameters and architecture around the optimal values for CNN, where each CNN as the base classifier was trained to output a probability for each class. Ashir, Abubakar M., and Alaa Eleyan [69] have proposed a new approach for improved facial expression recognition. The new approach is inspired by the compressive sensing theory and multi resolution approach to facial expression problems. Ghimire, Deepak et al. [70] have proposed a new method for the recognition of facial expressions from single image frame that uses combination of appearance and geometric features with support vector machines classification. In this paper the authors extracted region specific appearance features by dividing the whole face region into domain specific local regions. Geometric features are also extracted from corresponding domain specific regions. Qayyum, Huma, et al [71] have used stationary wavelet transform to extract features for facial expression recognition due to its good localization characteristics, in both spectral and spatial domains. More specifically a combination of horizontal and vertical subbands of stationary wavelet transform is used as these subbands contain muscle movement information for majority of the facial expressions. Feature dimensionality is further reduced by applying discrete cosine transform on these subbands. The selected features are then passed into feed forward neural network that is trained through back propagation algorithm. Mlakar, Uroš, et al [72] have proposed an efficient feature selection system applied to a Facial Expression Recognition (FER) system. This system, capable of recognizing seven prototypical emotions including neutral expression, is based on a histogram of oriented gradient descriptor (HOG) and difference feature vectors.

III. SURVEY FINDINGS

- Face and Expression Recognition are two main challenges in pattern recognition and computer vision.

- Physical appearance of face can have severe variations occurred due to variation in facial expression, illumination, occlusion, head pose, and aging.
- Most of the feature extraction algorithms developed and evaluated against the facial appearance and illumination variations. However, most of the algorithms have attained good recognition results only with more number of the training samples

IV. Research Direction

In recent years, the analysis of human affective behavior has been a point of attraction for many researchers. Such automatic analysis is useful in various fields such as psychology, computer science, linguistics, neuroscience etc. Such affective computing is responsible for developing standard systems and devices, useful for recognition and interpretation of various human faces and gestures. The emotions are categories as anger, disgust, fear, happiness, sadness and surprise. Such emotion recognition system involves three main steps: face detection, feature extraction and facial expression classification. Hence, there is a need for standard approaches that solve the problem of machines understanding the human affect behavior. Machine Learning approaches should be introduced to recognize the human emotion behavior.

V. Conclusion

Recognizing faces is a normal task for people, who usually do effortlessly and without much attention, but for machine learning in the area of Biometrics, it has remained a tricky problem. Now almost about 30 years of research in this area is just beginning to come out with helpful solutions and products. In biometrics research, automated face detection has a number of advantageous properties, which are translating research outcomes into realistic methods. This literature survey talked about the different views of the various researchers in the field of Facial Emotion Recognition using Image Processing techniques

References

[1] W. James, What is emotion? , Mind 9, 188-205, 1884
[2] N.L. Etcoff, J.J. Magee, Categorical perception of facial expressions, Cognition, 44, 227- 240, 1992
[3] Ekman, Paul. "Facial action coding system (FACS)." *A human face* (2002).
[4] Massaro, Dominic W., et al. "Picture my voice: Audio to visual speech synthesis using artificial neural networks." *AVSP'99-International Conference on Auditory-Visual Speech Processing*. 1999.
[5] Yamamoto, Eli, Satoshi Nakamura, and Kiyohiro Shikano. "Lip movement synthesis from speech based on Hidden Markov Models1." *Speech Communication* 26.1-2 (1998): 105-115.
[6] Tekalp, A. Murat, and Joern Ostermann. "Face and 2-D mesh animation in MPEG-4." *Signal Processing: Image Communication* 15.4-5 (2000): 387-421.
[7] Lyons, Michael, et al. "Coding facial expressions with gabor wavelets." *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*. IEEE, 1998.
[8] Calder, Andrew J., et al. "A principal component analysis of facial expressions." *Vision research* 41.9 (2001): 1179-1208.
[9] http://www.cs.bham.ac.uk/%7Eaxs/cogaff.html
[10] Kobayashi, Hiroshi, and Fumio Hara. "Recognition of six basic facial expression and their strength by neural network." *Robot and Human Communication, 1992. Proceedings., IEEE International Workshop on*. IEEE, 1992.
[11] Bregler, Christoph, Michele Covell, and Malcolm Slaney. "Video rewrite: Driving visual speech with audio." *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1997.
[12] Cosatto, Eric, Gerasimos Potamianos, and Hans Peter Graf. "Audio-Visual Unit Selection for the Synthesis of Photo-Realistic Talking-Heads." *IEEE International Conference on Multimedia and Expo (II)*. 2000.
[13] Ezzat, Tony, and Tomaso Poggio. "Miketalk: A talking facial display based on morphing visemes." *Computer Animation 98. Proceedings*. IEEE, 1998.
[14] Scherer, Klaus R. "Vocal affect expression: A review and a model for future research." *Psychological bulletin* 99.2 (1986): 143.
[15] Gutierrez-Osuna, Ricardo, et al. "Speech-driven facial animation with realistic dynamics." *IEEE transactions on multimedia* 7.1 (2005): 33-42.
[16] Hong, Pengyu, Zhen Wen, and Thomas S. Huang. "Real-time speech-driven face animation with expressions using neural networks." *IEEE Transactions on neural networks* 13.4 (2002): 916-927.
[17] Ashish Verma, L. Venkata Subramaniam, Nitendra Rajput, Chalapathy Neti, Tanveer A. Faruquie. Animating Expressive Faces Across Languages. IEEE Trans on Multimedia, Vol. 6, No. 6, Dec, 2004.
[18] Pavlovic, Vladimir I., Rajeev Sharma, and Thomas S. Huang. "Visual interpretation of hand gestures for human-computer interaction: A review." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 7 (1997): 677-695.
[19] J. K. Aggarwal, Q. Cai, Human Motion Analysis: A Review, Computer Vision and Image Understanding, Vol. 73, No. 3, 1999.
[20] D. M. Gavrila, The Visual Analysis of Human Movement: A Survey, Computer Vision and Image Understanding, Vol. 73, No.1. January, 1999: 82-98
[21] O'Brien, James F., et al. *Automatic joint parameter estimation from magnetic motion capture data*. Georgia Institute of Technology, 1999.
[22] A.Azarbayejani, etc., Real-Time 3-D Tracking of the Human Body, IMAGE'COM 96, Bordeaux, France, May 1996.
[23] Liu, Mengyi, et al. "Partial least squares regression on grassmannian manifold for emotion recognition." *Proceedings of the 15th ACM on International conference on multimodal interaction*. ACM, 2013.

[24] Ghimire, Deepak, and Joonwhoan Lee. "Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines." *Sensors*13.6 (2013): 7714-7734.
[25] Kahou, Samira Ebrahimi, et al. "Combining modality specific deep neural networks for emotion recognition in video." *Proceedings of the 15th ACM on International conference on multimodal interaction*. ACM, 2013.
[26] Zavaschi, Thiago HH, et al. "Fusion of feature sets and classifiers for facial expression recognition." *Expert Systems with Applications* 40.2 (2013): 646-655.
[27] Danisman, Taner, et al. "Intelligent pixels of interest selection with application to facial expression recognition using multilayer perceptron." *Signal Processing* 93.6 (2013): 1547-1556.
[28] Kalita, Jeemoni, and Karen Das. "Recognition of facial expression using eigenvector based distributed features and euclidean distance based decision making technique." *arXiv preprint arXiv:1303.0635* (2013).
[29] Wagner, Jennifer B., et al. "Eye-tracking, autonomic, and electrophysiological correlates of emotional face processing in adolescents with autism spectrum disorder." *Journal of autism and developmental disorders* 43.1 (2013): 188-199.
[30] Zhang, Li, et al. "Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot." *Expert Systems with Applications* 40.13 (2013): 5160-5168.
[31] Rivera, Adin Ramirez, Jorge Rojas Castillo, and Oksam Oksam Chae. "Local directional number pattern for face analysis: Face and expression recognition." *IEEE transactions on image processing* 22.5 (2013): 1740-1752.
[32] Sikka, Karan, et al. "Multiple kernel learning for emotion recognition in the wild." *Proceedings of the 15th ACM on International conference on multimodal interaction*. ACM, 2013.
[33] Ahsan, Tanveer, Taskeed Jabid, and Ui-Pil Chong. "Facial expression recognition using local transitional pattern on Gabor filtered facial images." *IETE Technical Review* 30.1 (2013): 47-52.
[34] Sadeghi, Hamid, Abolghasem-A. Raie, and Mohammad-Reza Mohammadi. "Facial expression recognition using geometric normalization and appearance representation." *Machine Vision and Image Processing (MVIP), 2013 8th Iranian Conference on*. IEEE, 2013.
[35] Cid, Felipe, et al. "A real time and robust facial expression recognition and imitation approach for affective human-robot interaction using gabor filtering." *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013.
[36] Young, Joanna Cahall, and Cathy Spatz Widom. "Long-term effects of child abuse and neglect on emotion processing in adulthood." *Child abuse & neglect* 38.8 (2014): 1369-1381.
[37] Majumder, Anima, Laxmidhar Behera, and Venkatesh K. Subramanian. "Emotion recognition from geometric facial features using self-organizing map." *Pattern Recognition* 47.3 (2014): 1282-1293.
[38] Owusu, Ebenezer, Yongzhao Zhan, and Qi Rong Mao. "A neural-AdaBoost based facial expression recognition system." *Expert Systems with Applications* 41.7 (2014): 3383-3390.
[39] Zhang, Ligang, Dian Tjondronegoro, and Vinod Chandran. "Random Gabor based templates for facial expression recognition in images with facial occlusion." *Neurocomputing*145 (2014): 451-464.
[40] Bejani, Mahdi, Davood Gharavian, and Nasrollah Moghaddam Charkari. "Audiovisual emotion recognition using ANOVA feature selection method and multi-classifier neural networks." *Neural Computing and Applications* 24.2 (2014): 399-412.
[41] Owusu, Ebenezer, Yonzhao Zhan, and Qi Rong Mao. "An SVM-AdaBoost facial expression recognition system." *Applied intelligence* 40.3 (2014): 536-545.
[42] Vitale, Jonathan, et al. "Affective facial expression processing via simulation: A probabilistic model." *Biologically Inspired Cognitive Architectures* 10 (2014): 30-41.
[43] Ali, Hasimah, et al. "Facial emotion recognition using empirical mode decomposition." *Expert Systems with Applications* 42.3 (2015): 1261-1277.
[44] Rice, Linda Marie, et al. "Computer-assisted face processing instruction improves emotion recognition, mentalizing, and social skills in students with ASD." *Journal of autism and developmental disorders* 45.7 (2015): 2176-2186.
[45] Hossain, M. Shamim, and Ghulam Muhammad. "Cloud-assisted speech and face recognition framework for health monitoring." *Mobile Networks and Applications* 20.3 (2015): 391-399.
[46] Ming, Yue. "Robust regional bounding spherical descriptor for 3D face recognition and emotion analysis." *Image and Vision Computing* 35 (2015): 14-22.
[47] Palo, H. K., Mihir Narayana Mohanty, and Mahesh Chandra. "Use of different features for emotion recognition using MLP network." *Computational Vision and Robotics*. Springer, New Delhi, 2015. 7-15.
[48] Barros, Pablo, et al. "Multimodal emotional state recognition using sequence-dependent deep hierarchical features." *Neural Networks* 72 (2015): 140-151.
[49] Zhang, Li, et al. "Adaptive facial point detection and emotion recognition for a humanoid robot." *Computer Vision and Image Understanding* 140 (2015): 93-114.
[50] Neoh, Siew Chin, et al. "Intelligent facial emotion recognition using a layered encoding cascade optimization model." *Applied Soft Computing* 34 (2015): 72-93.
[51] Mlakar, Uroš, and Božidar Potočnik. "Automated facial expression recognition based on histograms of oriented gradient feature vector differences." *Signal, Image and Video Processing* 9.1 (2015): 245-253.
[52] Liu, Mengyi, et al. "Au-inspired deep networks for facial expression feature learning." *Neurocomputing* 159 (2015): 126-136.
[53] Boughrara, Hayet, et al. "Facial expression recognition based on a mlp neural network using constructive training algorithm." *Multimedia Tools and Applications* 75.2 (2016): 709-731.
[54] Kahou, Samira Ebrahimi, et al. "Emonets: Multimodal deep learning approaches for emotion recognition in video." *Journal on Multimodal User Interfaces* 10.2 (2016): 99-111.
[55] Mano, Leandro Y., et al. "Exploiting IoT technologies for enhancing Health Smart Homes through patient identification and emotion recognition." *Computer Communications* 89 (2016): 178-190.
[56] Alhussein, Musaed. "Automatic facial emotion recognition using weber local descriptor for e-Healthcare system." *Cluster Computing* 19.1 (2016): 99-108.
[57] Kim, Bo-Kyeong, et al. "Hierarchical committee of deep convolutional neural networks for robust facial expression recognition." *Journal on Multimodal User Interfaces* 10.2 (2016): 173-189.
[58] Sun, Bo, et al. "Facial expression recognition in the wild based on multimodal texture features." *Journal of Electronic Imaging*25.6 (2016): 061407.
[59] Calvo, Manuel G., and Lauri Nummenmaa. "Perceptual and affective mechanisms in facial expression recognition: An integrative review." *Cognition and Emotion* 30.6 (2016): 1081-1106.
[60] Zhang, Li, et al. "Intelligent facial emotion recognition using moth-firefly optimization." *Knowledge-Based Systems* 111 (2016): 248-267.
[61] Zhang, Xiao, and Mohammad H. Mahoor. "Task-dependent multi-task multiple kernel learning for facial action unit detection." *Pattern Recognition* 51 (2016): 187-196.
[62] Altamura, Mario, et al. "Facial emotion recognition in bipolar disorder and healthy aging." *The Journal of nervous and mental disease* 204.3 (2016): 188-193.

[63] Kar, Nikunja Bihari, Korra Sathya Babu, and Sanjay Kumar Jena. "Face expression recognition using histograms of oriented gradients with reduced features." *Proceedings of International Conference on Computer Vision and Image Processing*. Springer, Singapore, 2017.

[64] Liu, Zhentao, et al. "A facial expression emotion recognition based human-robot interaction system." (2017).

[65] Lopes, André Teixeira, et al. "Facial expression recognition with convolutional neural networks: coping with few data and the training sample order." *Pattern Recognition* 61 (2017): 610-628.

[66] Fan, Xijian, and Tardi Tjahjadi. "A dynamic framework based on local Zernike moment and motion history image for facial expression recognition." *Pattern Recognition* 64 (2017): 399-406.

[67] Datta, Samyak, Debashis Sen, and R. Balasubramanian. "Integrating geometric and textural features for facial emotion classification using SVM frameworks." *Proceedings of International Conference on Computer Vision and Image Processing*. Springer, Singapore, 2017.

[68] Wen, Guihua, et al. "Ensemble of deep neural networks with probability-based fusion for facial expression recognition." *Cognitive Computation* 9.5 (2017): 597-610.

[69] Ashir, Abubakar M., and Alaa Eleyan. "Facial expression recognition based on image pyramid and single-branch decision tree." *Signal, Image and Video Processing* 11.6 (2017): 1017-1024.

[70] Ghimire, Deepak et al. "Facial expression recognition based on local region specific features and support vector machines." *Multimedia Tools and Applications* 76.6 (2017): 7803-7821.

[71] Qayyum, Huma, et al. "Facial expression recognition using stationary wavelet transform features." *Mathematical Problems in Engineering* 2017 (2017).

[72] Mlakar, Uroš, et al. "Multi-objective differential evolution for feature selection in facial expression recognition systems." *Expert Systems with Applications* 89 (2017): 129-137.