The 7th IEEE International Conference on E-Health and Bioengineering - EHB 2019
Grigore T. Popa University of Medicine and Pharmacy, Iasi, Romania, November 21-23, 2019

Analysis on Emotion Detection and Recognition Methods using Facial Microexpressions. A Review

Loredana Stanciu[1], Adriana Albu[1]
[1] Automation and Applied Informatics Department, Politehnica University Timisoara, Timisoara, Romania, loredana.stanciu@upt.ro, adriana.albu@upt.ro

Abstract— **Facial recognition is already an integral part of many commercial applications and, of course, the recognition of emotions based on facial expressions may become as usual soon. It is currently used in applications in the field of psychotherapy, security systems, forensics, etc. The face is the primary means of communicating emotions, but it is also the one that hides them most easily. Micro-expressions are a very accurate indicator of our subtle and involuntary thoughts, showing what we really feel. Unfortunately, their duration and intensity are very short, which makes their detection very difficult, even trained people can detect below 50%. For this reason, automatic detection becomes more interesting and challenging, and research in this direction is in full development. The article contributes to this area by presenting the general process of automatic detection and synthesizing the latest techniques in classification and recognition of facial micro-expressions, which, although diverse, have an accuracy of recognition around 80%.**

Keywords—affective computing; Computer Vision; face detection; micro-expression.

I. INTRODUCTION

We are born with the facial expressions of six emotions: disgust, fear, happiness, surprise, sadness, anger, which are deeply imprinted in our bodies (Fig. 1). Lately, the scientist added an extra emotion, which is contempt. Just like a preloaded language in our bodies database, these seven universal emotions are effortlessly recognized by our mirror-neuron systems in every human interaction we go through. The face is the primary means of communicating emotions, but it is also the one that hides them most easily. When we express our emotions openly, they are easy to perceive. Even if the expressions remain on the face for a period ranging from 1/25 fractions of a second to 4 seconds, our brains can process the information received even in such a short time. On the other hand, capturing true emotion on a movie is a real challenge, especially in today's world, where there is a phobia for feelings and most people don't leave the house unless they put on a mask [1].

Fig. 1. Example of universal emotions. From left to right: disgust, fear, happiness, surprise, sadness, and anger [14]

Today's most important role in influencing contemporary life is undoubtedly technology, and automation and computerization are topics that are increasingly being discussed both in the scientific community and in other fields. Technological advancement leads to cheaper equipment and increased software power. In this context, also the need to find a way for humans to interact with computers appeared and the computers to respond to people according to their needs. The domain in which these things are possible is called affective computing, an interdisciplinary field that covers several sciences such as: computer science, psychology and cognitive sciences. This is a form of technology based on which the computer is able to respond in some way to a human stimulus, usually one that is related to mood or emotions - whether it interprets human gestures and body language, measures vital signs or other information to generally understand and respond to the user's mood, emotional or general status.

Detecting emotions using micro-expressions can be an advantage for the security of citizens. Thus, an interrogation would be much more efficient and faster in obtaining information, the lies would be much easier to identify, being an important leap from the classic polygraph, where there is evidence that a proper preparation can deceive it [2]. Also, this technology can help to better understand the nervous system and some diseases. Research has shown that people suffering from Huntingston's disease have a deficiency in emotion recognition. They have very little accuracy in detecting facial expressions that express anger and disgust and a mean one for sadness and surprise, to which is added a deficit in recognizing the intensity of surprise and disgust [3].

This paper aims to present a quick survey of the latest methods for detecting and interpreting facial micro-expressions that identify, as accurately as possible, the expressed emotion. The subject is a difficult one to approach because there is not yet a standardized procedure but several procedures, some of

978-1-7281-2603-6/19/$31.00 ©2019 IEEE

them more difficult to use than others and which require certain conditions for optimal functioning.

II. Background

Facial expression is an important form of expression of the emotional and mental state, the information transmitted in this way reaching 55% of the total communication [4, 5, 6]. For this reason, facial expression recognition has become an important research topic. The automatic analysis of facial expressions contains three steps: acquiring the face image, extracting and representing the data of interest and recognizing the expression. There are two approaches to data representation: feature-based geometric approach and appearance-based approach [7, 8]. In the feature-based geometric approach, facial points (lip corners, eye center, eyebrow edges and nose tip) are extracted using image processing techniques and the coordinates obtained create a characteristic vector, representing facial geometry. Appearance-based methods analyze video frame by frame and use the image filter to extract a feature vector. This can be applied on the full face or on a certain region.

Computerized analysis of facial expressions has several limitations. Expressions contain information about both face identity and expression changes, so the extracted features are often a mixture of expression changes and identity features. The way of illuminating the face from different directions and different intensities has a significant impact on the feature extraction. If key areas, such as the eyes or mouth, are dimly lit, this will affect feature extraction and interfere with facial expression recognition [4].

Psychologists classify these expressions in posed or spontaneous, different from each other in terms of appearance and temporal characteristics [7]. Automated facial expression recognition systems have focused on posed expressions, while our daily interactions have spontaneous facial expressions, which greatly reduces the accuracy of recognition of these automated systems. Another limitation is that micro-expressions have a very limited duration and low intensity and automatic recognition requires a sufficiently large and varied database regarding recorded emotions [4, 9].

Another interesting aspect is suggested by neuropsychological studies: the two parts of the human face are not equal pronounced during emotional expressions, being more intense on the left side of the face. The studies of facial asymmetry also show that socially appropriate cues are more pronounced on the right side of the face, while personalized feelings are visible on the left side of the face, evidence that anatomically supports

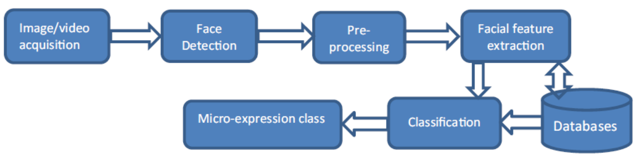

Fig. 2. The process of micro-expression recognition [15]

the difference between posed and spontaneous expressions [7].

There are numerous studies in psychology that suggest the universality of at least six basic emotions - happiness, sadness, fear, anger, surprise and disgust [3]. Facial expressions differ from facial movement, recognizing them is much easier when comparing the neutral state with any other emotional expressions. Moreover, the faces have permanent features and transient features. Eyes and lips act as permanent features, while facial lines and expression wrinkles act as transient features. Studies show that expressions of surprise, fear, disgust and anger produce more facial movement and suggest that the upper and lower halves should be analyzed separately [7].

Also, there are expressions that are detected more easily (for example, sadness can only be recognized by looking at the mouth), and more difficult ones, even confused with others (anger with disgust and fear with surprise), probably because they contain common facial movements. Also, emotions such as surprise, fear, disgust and anger, produce more facial movements compared to sadness and smile. In addition, there are two perspectives on how to make recognition [7]: one perspective argues for component recognition, while the other argues that recognition is a holistic process.

Micro-expression is a short facial movement that reveals an emotion a person is trying to hide [2, 9] or, very often, the person in question is not aware of. There are trained experts who can identify the existence and recognize micro-expressions, but the accuracy is only about 47% [10]. For this reason, it can reveal the true feelings of the person or an attempt of lie. Micro-expressions have three significant characteristics [2]: short duration, low intensity and they are fragments of prototypical facial expressions.

Facial micro-expressions were discovered by Haggard and Isaacs in 1966 as they looked over records from psychotherapists and searched for non-verbal language cues between patient and therapist [2, 20]. Then, Ekman and Friesen in 1969 and 1974 include the concept of recognizing micro-expressions in their study of deception. The result was published in the book Telling Lies (Ekman, 1985) and was popularized in the series Lie to Me, the first scientific article on micro-expressions appearing much later, in 2000 [3, 11].

The Face Action Coding System (FACS) is a taxonomy system of human facial movements, based on a system originally developed by Carl-Herman and updated by Ekman, Friesen and Joseph [11]. The development of FACS with 44 Action Units (AU) to describe the facial expressions [5] was one of the most significant steps toward understanding them. The seven basic facial actions related to emotion can be represented by the following areas on the face: the left and right edges of both eyebrows, the left and right edges of the eyes, the left and right edges of the mouth and cheeks [11].

Facial emotion recognition methods could be categorized as image-based, video-based, and 3D surface-based [5]. The process of micro-expression recognition consists in some specific steps, as shown in Fig. 2, the most important being face detection, pre-processing, facial feature extraction, classification and recognition. There are many attempts to make this process faster and more accurate, some of the latest being presented in the next section.

A great limitation in identifying emotions based on micro-expressions, most of the researchers face, is the lack of a standard micro-expression database, which makes very difficult the training of an accurate micro-expression recognition system [10]. The most used databased are Spontaneous Micro-Expression Database (SMIC), The Chinese Academy of Sciences Micro-Expression (CASME and CASME II), Micro-Expression Training Toolkit (METT), Unsupervised Segmentation Fusion-High Definition (USF-HD), York Deception Detection Test (York DDT).

III. LATEST STUDIES

Because tracking each significant trait point for micro-expression detection requires a lot of data, Yao et al [11] decided to choose specific operating points and associated Action Units (AU) like lip edges, because they are easier to detect and track. To improve the accuracy of recognition, they have merged a Local Binary Pattern (LBP) and a pseudo 3D model to obtain the face area and its different regions. Then, these facial regions are used to train Hough Forest (HF) to extract the characteristic points. During the test procedure, the first image frame is divided into several regions, and the buccal region would be tested by HF for lip delineation. A Tracking Learning Detection algorithm receives the position of the lip edges and can learn and follow the characteristic points by itself. The method is suitable for emotions in which the position of the mouth is very important. The emotions detected were happiness (84%) and disgust (74.5%).

Also based on facial points, but 3D, Zhang et al. [5] propose to identify three different types of human emotions: sad, happy and neutral. Kinect V2.0 is used to record the change of facial expression of the person over a period of time. For each frame, 1347 facial 3D points are captured from which 100 key facial points are selected. Then, Moving Middle Filter is used to eliminate noise and the data flow is divided into several slices with uniform dimensions, considered samples, from which efficient attributes are extracted. The recognition is done with several classic classifiers with default parameter based on the Weka software, and the recognition rates are for both men and women, 70%, 77% and 80%, respectively.

Wang et al. [2] propose a different method, a color space model, Tensor Independent Color Space (TICS). A color video with micro-expressions is treated as a four-order tensor, that is, a four-dimensional array. The first two dimensions are spatial information, the third is temporal information, and the fourth is color information. They turn the fourth dimension from RGB into TICS, where the color components are as independent as possible. They also define a set of Regions of Interest (ROI) based on the FACS coding system and calculate dynamic texture histograms for each ROI. The experiments are performed on CASME and CASME 2, and the results show that the performance for TICS is better than for RGB or gray, the accuracy being between 50% and 72.09%.

Davison et al. [12] define a face template consisting of 26 regions based on FACS and then extract the temporal features of each region using 3D Histogram of Oriented Gradients (3D HOG). They search for subtle facial motion in the local regions with Chi-square distance. Finally, an automatic peak detector is used to detect micro-movements above un adaptive baseline threshold. The results show that 3D HOG outperformed for micro-movement detection, compared to Local Binary Patterns in Three Orthogonal Planes (LBP-TOP) and Histograms of Oriented Optical Flow (HOOF) on SAMM and CASME II, with an accuracy of 73.55%.

Davison and his team [13] came with the idea of modifying the classification of expressions from the CASME II database on the basis of AUs and not on the predicted emotions, in order to remove the potential bias of human reporting. They obtained 5 classes which were tested using LBP-TOP, HOOF and HOG 3D feature descriptors on two databases CASME II and SAMM (A Spontaneous Micro-Facial Movement). The best accuracy was obtained in the combination of HOG 3D and CASME II, 86.35%.

A face dynamic map (FDM) is proposed by Xu et al. [14], for differentiating images with a micro-expression from those without micro-expression (or non-micro-expression) and for classifying different micro-expressions, especially in situations when they are spontaneous. An algorithm based on optical flux estimation is used to perform pixel-level alignment for micro-expression sequences. Then different levels of facial dynamics are determined, in different granularities based on spatio-temporal cuboids, and a classifier, such as Support Vector Machines (SVM), is used for both identification and classification. The experimental validation was done on several sets of reference micro-expression data, the recognition rates are between 50-70%, depending on the database used.

A visual platform is presented in [8] for the recognition of facial micro-expressions. The platform uses the Gabor wavelet filter to extract expression characteristics, principal component analysis (PCA) and linear discriminant analysis (LDA) for size reduction and SVM with Haar Cascades for expression classification. The platform visualizes the results obtained after each step in processing. The experimental results show that the proposed scheme works well with the CASME II database. Recognition rates are very high, even 100% for still images, but much lower for video recognition. Also Haar Cascades Classifier for facial recognition is used in [15]. The proposed method was implemented using OpenCV, an open source Computer Vision and machine learning software library, having as detection algorithm the Viola-Jones method. The accuracy is up to 90% for happiness and sadness in static images.

In [16] the authors focus on the analysis of spontaneous facial micro-expressions. They propose a Spatio-Temporal Completed Local Quantized Pattern (STCLQP) method that exploits three information: sign-based, magnitude-based and orientation-based difference of pixels. They use vector quantization and codebook selection for each component in appearance and temporal domains to learn compact and discriminative codebooks for generalizing classical pattern types. Based on those, spatiotemporal features of sign, magnitude and orientation components are extracted and fused. The results showed that all three components can improve the performance of spontaneous micro-expression analysis, but the orientation component can perform better.

The authors of [17] use a deep multi-task learning method to train a deep convolutional neural networks (CNN) for facial landmark localization, crucial in micro-expression analyzing. The facial area is split into 12 ROIS. Then they combine robust optical flow approach with the HOOF feature for evaluating the direction of movement of facial muscles. As a classifier for detecting micro-expression they used SVM. The detection accuracy on CASME database is 80%.

Also CNN are used in [18]. To implement the micro-expressions detection network, they first trained the network on selected frames from the training data sequences extracted from CASME II. Then they combined the convolutional layers from the trained network, with a long-short-tern-memory recurrent neural network, whose input is connected to the first fully connected layer of the feature extractor CNN. The experiments revealed a detection accuracy of 59.47% better than in other methods (LBP-TOP and LBP-TOP with adaptive magnification).

A model pre-trained CNN was used in [19] for the classification of micro-expressions. The results showed that in order to obtain the best possible recognition accuracy, it is crucial that the image databases for training be as large as possible. As a result, they combined the two CASME and CASME II databases to form a larger database. The accuracy obtained in this way was 75.57% on CASME II.

IV. RESULTS AND CONCLUSIONS

There are several ways for face detection: specific operating points [11] or 3D points [5], color space model [2], face template based on FACS [12], dynamic map [14], Gabor wavelet filter [8], vector quantization and codebook selection [16] or CNNs [17, 18, 19]. Also, for emotion detection one can use Tracking Learning Detection algorithms [11], different kinds of classifiers [5, 8, 14, 15, 17], histograms [2, 12, 13, 17] or neural networks [18, 19]. Although different modalities have been used for facial detection and emotion recognition, the accuracy of recognition is in the range of 70-85%, higher only in specific situations [13] and for certain types of emotions, easier to detect [8, 15]. These automated facial expression recognition systems have focused on posed expressions, more than spontaneous, which actually occur in our daily interactions. This is explainable because is much harder for the latter to be detected.

The analysis performed on the methods of detecting emotions using facial micro-expressions gives us the answer that we kind of expected: there is no perfect method for achieving this goal. The volume of computer techniques applied even in the case of methods that seem "easy" is particularly high. The accuracy is also dependent on the quality of the images used to train the classifiers or neural networks, the existing databases being in some cases insufficient.

As the success rate in identifying emotion based on facial micro-expression increases, so does the complexity of the methods and the need for computing power to obtain the result, which causes us to wonder how much we are willing to invest in technology to achieve the highest accuracy? However, it is possible that the technological development of the following years will allow even earlier than we think the recognition of emotions integrated in commercial devices, as the facial recognition already is.

REFERENCES

[1] V. Bettadapura, “Face expression recognition and analysis: the state of the art”, ArXiv 2012.

[2] S.-J. Wang, W.-J. Yan, X. Li, G. Zhao, C.-G. Zhou et al., “Micro-expression recognition using color spaces”, IEEE Transactions on Image Processing, **24(12)**, 2015, pp. 6034–6047.

[3] D. Matsumoto, J. LeRoux, C. Wilson-Cohn et al., “A new test to measure emotion recognition ability: Matsumoto and Ekman’s Japanese and Caucasian Brief Affect Recognition Test (JACBART)”, Journal of Nonverbal Behavior, **24(3)**, 2000, pp. 179-207.

[4] T. Wu, S. Fu, G. Yang, “Survey of the facial expression recognition research”, International Conference on Brain Inspired Cognitive Systems, 2012, pp. 392–402.

[5] Z. Zhang, L. Cui, X. Liu, T. Zhu, “Emotion detection using Kinect 3D facial points”. IEEE/WIC/ACM International Conference on Web Intelligence (WI), 2016.

[6] R. Sharma, B. Kaushik, “Facial expression recognition: a survey”, International Journal of Computer Applications, 153, 2016, pp. 32-36.

[7] B. Bhushan, “Study of facial micro-expressions in psychology”, Understanding Facial Expressions in Communication, pp. 265–286, 2014.

[8] P. Zhang, X. Ben, R. Yan, C. Wu, C. Guo, “Micro-expression recognition system”, Optik - International Journal for Light and Electron Optics, **127(3)**, 2016, pp. 1395–1400.

[9] W.-J. Yan, S.-J. Wang, Y.-J. Liu, Q. Wu, X. Fu, (2014). “For micro-expression recognition: database and suggestions”, Neurocomputing, **136**, 2014, pp. 82–87.

[10] M. Takalkar, M. Xu, Q. Wu, Z. Chaczko, “A survey: facial micro-expression recognition”. Multimedia Tools and Applications, **77(15)**, 2017, pp. 19301–19325.

[11] S. Yao, N. He, H. Zhang, O. Yoshie, “Micro-expression recognition by feature points tracking”, 10th International Conference on Communications (COMM), 2014.

[12] A. Davison, W. Merghani, C. Lansley, C.-C. Ng, M.H. Yap, “Objective micro-facial movement detection using FACS-based regions and baseline evaluation”. 13th IEEE International Conference on Automatic Face & Gesture Recognition, 2018, pp.642-649.

[13] A. Davison, W. Merghani, M.H. Yap, “Objective classes for micro-facial expression recognition”, Journal of imaging, **4**, 2018.

[14] F. Xu, J. Zhang, J. Z. Wang, “Microexpression identification and categorization using a facial dynamics map”, IEEE Transactions on Affective Computing, **8(2)**, 2017, pp. 254–267.

[15] L. Stanciu, F. Blidariu, “Emotional states recognition by interpreting facial features”, E-Health and Bioengineering Conference (EHB), 2017, pp. 273-276.

[16] X. Huang, G. Zhao, X. Hong, W. Zheng, M. Pietikäinen, “Spontaneous facial micro-expression analysis using Spatiotemporal Completed Local Quantized Patterns”. Neurocomputing, **175**, 2016, pp. 564–578.

[17] X. Li, J. Yu, S. Zhan, “Spontaneous facial micro-expression detection based on deep learning”. IEEE 13th International Conference on Signal Processing (ICSP), 2016, pp. 1130-1134.

[18] R. Breuer, R. Kimmel, “A deep learning perspective on the origin of facial expressions”, Computer Vision and Pattern Recognition, ArXiv 2017.

[19] M. Takalkar, M. Xu, “Image based facial micro-expression recognition using deep learning on small datasets”, International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2017

[20] H. Costin, C. Rotariu, I. Alexa, et al., “TELEMON - A complex system for real time medical telemonitoring”, 11th Int. Congress of the IUPESM/World Congress on Medical Physics and Biomedical Engineering, Munich, Germany, Sept. 07-12, 2009.