SEUSL

Sri Lankan Journal of Technology

Research

Open Access

SLJoT
2021 Sp Issue 01-10
ISSN 2773-6970

A Study of Facial Emotion Recognition Techniques to Examine Micro-Expressions

Dewmini A.G.H.U.[1*], Hirshan R.[2] and Kumara W.G.C.W[3]

[1,2]Dept. of Elec. and Telecom. Engineering, Faculty of Engineering, South Eastern University of Sri Lanka
[3]Dept. of Computer Science and Engineering, Faculty of Engineering, South Eastern University of Sri Lanka
[*]Corresponding Author: h.u.dewmi@gmail.com

Received: 25-04-2021 ∗ Accepted: 14-07-2021 ∗ Published Online: 15-09-2021

***Abstract*—**Humans communicate with one another by speaking, gesticulating with their bodies, and expressing facial emotions. Among these methods, expressing emotions play an important role. Since human beings naturally use facial expressions to convey their emotions. Micro-expressions are perceptive facial expressions that last only a few seconds. Micro-expressions, as compared to regular facial expressions, will expose the majority of the latent, unconcealed emotional states. However, because of their shorter length, micro-expressions are more difficult to find. As a result, interest in micro-expression has grown in many fields, including defence, psychology, and computer vision, in recent years. This paper provides a brief overview of current methodologies for detecting human micro-emotions, with a focus on the LBP, LBP-TOP, DCNN, 3DHOG, MMPTR, and DTCM feature extraction filter methods, which have been found to be more accurate. The theoretical accuracy of the LBP-TOP Feature Extraction method with SVM and KNN classifier combination was discovered to be better than the theoretical accuracy of all approaches. As a result, this paper also discusses those two classifiers.

***Keywords*—**Micro-expressions, feature extraction techniques, spontaneous datasets, Eulerian Video Magnification

I. Introduction

Ekman and Friesen have identified seven universal facial expressions based on extensive research: happiness, surprise, fear, disgust, sadness, contempt and anger. It is possible to recognize one or more of these emotional classes in a single video frame while filming an emotional event. (Merghani *et al.*, 2018).

Normal facial expressions (macro-expressions) are visible, but micro-expressions are so much faster and effort to suppress the original emotions while concealing the exact emotion felt. Because of the rapid change of facial expressions, it is difficult to recognize a person's true emotions in a video stream using human senses. Ekman *et al*. (Wu *et al*., 2011) created a training algorithm that is used to train emotion databases to distinguish micro-expressions effectually in order to investigate these involuntary expressions. They once examined a video of a psychiatric patient who later confessed to lying to her doctor about her intention to commit suicide (Wu *et al*., 2011). Although the patient appeared to be happy throughout the video series, when the video frames were examined one by one, a concealed anguish expression was discovered that lasted for only two frames (which lasted approximately 0.08 seconds) (Wu *et al*., 2011). This indicates that it is prominent to recognize humans' micro-expressions.

According to previous research studies, two main types of facial feature extraction methods. They are appearance-based methods and geometric feature-based methods, respectively (Huang *et al*., 2016). Geometric methods are used to convert facial landmarks in face geometry into feature vectors. To track appearance changes, appearance-based methods are used. To identify the appearance changes, image filters are applied to either the entire face or specific face regions such as the nose and mouth.

A comprehensive examination of the data sets and methodologies for Micro-Expression recognition were presented in this study. The rest of the paper is structured as follows: Section II evaluates an investigation of existing datasets. Section III methodology discusses techniques for extracting features and emotion classification methods, Section IV discusses research challenges and limitations, and Section V concludes this paper.

II. FACIAL MICRO-EXPRESSION DATASET

Spontaneous data sets and non-spontaneous data sets are the two major types of Micro-Expression data sets. Non-spontaneous data sets differed significantly from spontaneous data sets because they were not collected spontaneously. Non-spontaneous data sets include the Polikovsky Dataset, the York Deception Detection Test (YorkDDT), and the USF-HD, and It was observed that they are not presently publicly available for study. As a result, the datasets that could only be considered were those that were created spontaneously.

Copyright ©2021 belongs to Faculty of Technology, South Eastern University of Sri Lanka, University Park, Oluvil, #32360, Sri Lanka

A. Spontaneous Datasets

To obtain an accurate data set, true emotions must be captured while the person is attempting to conceal them, rather than using fake emotions created by actors. Though eliciting Micro-Expressions spontaneously is difficult, there are some available spontaneous data sets, including; SMIC data set (Spontaneous Micro-expression Corpus), the CASME dataset (Chinese Academy of Sciences Micro-Expressions), (CASME II), CAS(ME)2 (A Data-set of Spontaneous Macro-Expressions and Micro-Expressions), and the SAMM dataset (Spontaneous Actions and Micro-Movements). Table I illustrates the characteristics of the abovementioned data sets. The researchers' attention has shifted to CASME II and SAMM because they are well-equipped with emotion classes, have a high frame rate, and a wide range of intensity of facial movements (Merghani, 2018). Table II demonstrates the data classes of each spontaneous data set for greater clarity.

Emotion classes in the SMIC data set have been classified as positive, negative, and surprise. As a result, that data set cannot be used to identify different levels of human emotion. On the other hand, the CASME data set series contains a large number of valuable datasets; for example, in the CASME data-set can identify, in addition to the basic emotions, there is Tenseness, and Repression. Most importantly, the CASME2 dataset contributes more to micro-expression recognition systems because it contains 250 micro-expressions.

B. CASME, CASME II, and CASME2 Dataset

Wen-Jing Yan, Q. Wu, Yong-Jin Liu, Su-Jing Wang, and X. Fu created the CASME dataset in 2013 (Yan *et al*., 2014). Its improved versions are CASME II and CASME2. They have provided a reasonably effective and efficient spontaneous micro-expression database, which includes samples of dynamic micro-expressions as well as spontaneous micro-expressions, Action Units (AU) to each micro-expression, and the emotion classes that occur have been prudently labelled based on psychological studies and participant self-report. Furthermore, the un-emotional facial movements have been detached from the data set. As a result, this is the best existing dataset that can be used for study in the field of micro-expression recognition. This dataset contains raw video clips of facial expressions as well as image frames of each video frame in .jpg format. If a researcher wants to analyse videos and frame them into their ratio, they can use given video clips. However, the dataset itself contains framed images that are labelled with the Action Units. Consequently, it is simpler to use. CASME II has a higher sampling rate than the other datasets, at 200 frames per second with the goal of providing more comprehensive information on facial muscle movements. Furthermore, the images in CASME II have a larger face dimensions; approximately 280 pixels x 340 pixels when compared to the face size of other data sets in CASME which is approximately 150 pixels x 190 pixels.

Moreover, unlike other datasets, CASME provides an adequate number of datasets for training and testing because

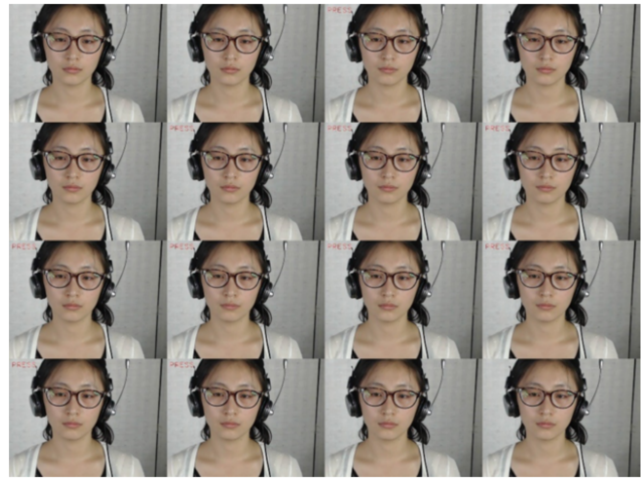

Figure 1: CASME data set 'Anger' Micro-expression video frames

Figure 2: Images with different type of eyeglasses

robust automatic Micro-Expression recognition explorations and associated applications necessitate a larger number of samples. The video episodes were recorded in various color variations to help the model understand emotions under different lighting conditions. CASME recorded datasets in natural light, a room with two LED lights, and four carefully selected LED lamps. Aside from that, they have video episodes with both male and female participants. In addition, they used some other background equipment such as eyeglasses and microphones with participants to provide variations to the dataset that did not only have plane human faces. This is also advantageous when developing the emotion recognition model, as it allows the model to be trained more precisely. Figure 01 is a screen capture from a video episode that has a frame rate of 30 frames/second which depicts a variety of angry emotion facial expressions.

C. Pre-processing

Detecting and tracking facial landmarks, face cataloging, face masking, and removing unwanted background are the general pre-processing steps in facial Micro-Expression spotting.

1) Using Action Units:: Using Action Units of Human Face is one method for identifying face regions and removing unwanted backgrounds. The CASME dataset contains a sufficient number of Action Unit (AU) coordination. If a researcher wants to use the labeled Action Unit details to detect the face region, he or she can use the provided Action Unit dataset. However, in some cases, coordination is

Copyright ©2021 belongs to Faculty of Technology, South Eastern University of Sri Lanka, University Park, Oluvil, #32360, Sri Lanka

Table I: Available sponpicttaneous datasets (Merghani, 2018)

Dataset	SMIC	CASME	CASME II	CAS(ME)2	SAMM
Number of participants	20	35	35	22	32
Frames per second	100 and 25	60	200	30	200
Resolution (pix x pix)	640 x 480	640 x 480, 1280 x 720	640 x 480	640 x 480	2040 x 1088
Number of data samples	164	195	247	53	159
Number of emotion classes	3	7	5	4	7
Is the FAC (Facial Action Coding) encoded?	NO	YES	YES	NO	YES

Table II: Emotion classes of each dataset (Merghani, 2018)

Dataset	Emotion Classes
SMIC	Positive, Surprise, Negative
CASME	Surprise, Amusement, Disgust, Sadness, Fear, Contempt, Tense, Repression
CASME II	Surprise, Happiness, Repression, Disgust, Others
CAS(ME)2	Positive, Surprise, Negative, Other
SAMM	Surprise, Happiness, Disgust, Sadness, Fear, Contempt, Anger

insufficient, because the human face region is generally identified using 30 Action Units. Ekman and Friesen identified 30 Action Units from the face region, two for the upper face and eighteen for the lower face. Action Units can occur separately or in combination. Figures 02 and 3 show some of the most common AUs as well as some additive and nonadditive AU combinations.

Action Units can be used to automatically identify face region, but they can also be used to identify facial expression type, as face muscle coordination changes with emotion level. High precision AU Recognition System methods that can be recommended include feature-based AU recognition, Optic-flow AU recognition, ICA or Gabor wavelet method, Dense-flow method, and Edge-detection method. These methods are capable of providing nearly 90 percent accuracy.

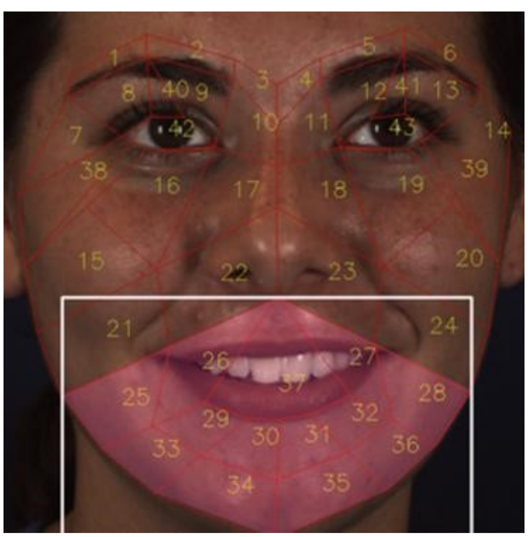

Figure 3: This image summarizes all Action Units(Tian *et al.*, 2001)

2) Using OpenCV: : Object detection using Haar feature-based cascade classifiers is an efficient method for detecting faces in static images. However, if detecting a face region in a video stream was expected, OpenCV's Haar cascades could be used to recognize faces in real-time video streams. Face detection with Haar cascades is a machine learning technique that comprises training a cascade function with a set of input data. OpenCV, which affords real-time optimized Computer Vision tools, hardware, and library already includes a significant amount of pre-trained classifiers for the face, emotions, and so on. However, in terms of speed, the Histogram of Oriented Gradients (HoG) algorithm appears to be the fastest, followed by the Haar Cascade classifier and Convolutional Neural Networks (CNNs).

3) Eulerian Video Magnification (EVM): : According to the definition of Micro-Expressions, those reactions occur in a fraction of a second. As a result, detecting those emotions with the naked eye is enormously challenging. Even though Micro-Expressions are recorded in video streams, they are difficult to detect with the naked human eye. In that context, the ability to amplify the original video to disclose concealed information in video streams by applying temporal filtering to the frames will be advantageous in detecting subtle micro-expressions. One of the new techniques for Micro-Expression preprocessing is Eulerian Video Magnification (EVM), which is used to amplify subtle facial expression movements. Motion magnification begins with the application of spatial decomposition pyramid and temporal filtering to video frames, followed by the amplification of the resulting signals to reveal hidden motion information. After that, the motion-amped face video sequence is used to extract Spatio-temporal feature patterns. The framework's main procedure is as follows: construct the Laplacian pyramid, perform temporal filtering on each level (and possibly some) of the Laplacian pyramid, perform pixel change magnification on the temporal filtering result; and reconstruct the image by changing the value of pixels on the Laplacian pyramid (Eulerian Video Magnification, 2021).

4) Total Variation: L1 Optical Flow Algorithm (TV-L1): Micro-expression videos can be pre-processed using the Total Variation-L1 (TV-L1) method for optical flow approximation. The TV-L1 has two major advantages: improved noise robustness and preservation of flow discontinuity. In the optical

Copyright ©2021 belongs to Faculty of Technology, South Eastern University of Sri Lanka, University Park, Oluvil, #32360, Sri Lanka

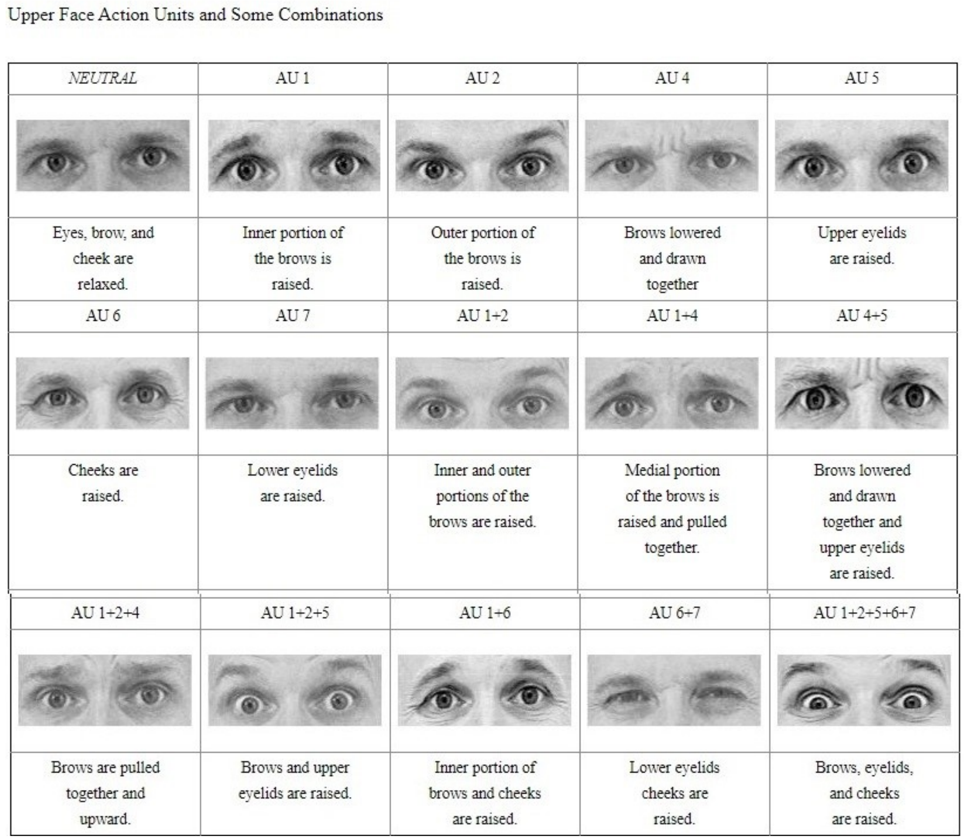

Figure 4: Sample Upper Face Action Units and Action Unit combinations (Papers with Code - Action Unit Detection, 2021

Figure 5: Face Cropped Anger Micro-Expression Video Frames

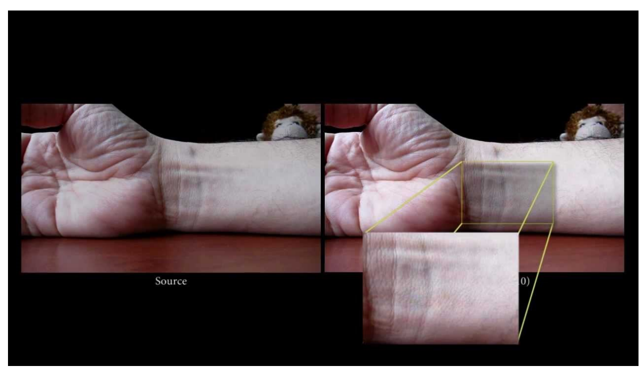

Figure 6: Before and After Eulerian Video Magnification (Eulerian Video Magnification, 2021)

flow algorithm, Total Variation-L1 (TV-L1) computes the motion information between the onset and apex frames. The conforming optical flow function map was then obtained, and deep network learning was performed. The algorithm keeps the image's discontinuity, as well as its edge features and other feature detail. To improve registration robustness, the optical flow employs the L1 norm (Zhao and Xu, 2019).

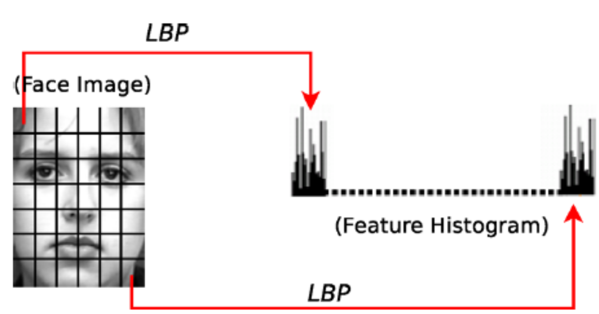

Figure 7: Face image is converted into a histogram (Shan *et al*., 2009)

Copyright ©2021 belongs to Faculty of Technology, South Eastern University of Sri Lanka, University Park, Oluvil, #32360, Sri Lanka

III. Methodology

A. Feature Extraction

This section discusses the most effective feature extraction methods and emotion classification techniques that have been proven to provide high accuracy.

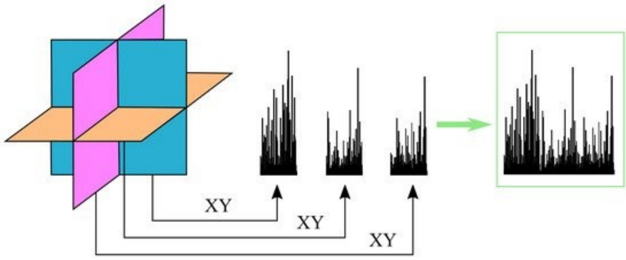

Figure 8: The Histogram of Concatenated LBP-TOP feature (Pfister *et al*., 2011)

1) Local Binary Pattern or LBP: The LBP operator marks pixels in an image by comparing the values of each pixel to the values of its neighbours. The results are in binary number format, and a texture descriptor was created using a 256-bin histogram of the LBP labels computed over an area. The LBP histogram comprehends statistics about the scattering of local micro-patterns such as flat areas, spots, and edges across the entire image and can thus be used to statistically define properties. For example, consider the frontal-face region. Figure 5 shows how the frontal face was divided into small regions to obtain LBP histograms. The extracted feature function histogram constitutes the local texture and global form of face images. To recognize faces using the LBP-based facial representation, first create a template for each emotion class of images, after that use a Nearest-Neighbour (NN) Classifier to align the input image with the closest template. In this case, pattern matching was utilized to characterize facial expressions for the sake of ease. During training, the histograms of expression images in that class were averaged to form a template for that particular class (Shan *et al*., 2009).

2) Local Binary Pattern-Three Orthogonal Planes or LBP-TOP: A video sequence can be represented as a stack of YT, XT, and XY planes with time T-axis, spatial X-axis, and spatial Y-axis. The XY plane primarily disclosures spatial data, whereas the XT and YT provide detailed information on how pixel grayscale values transit over time (Pfister *et al*., 2011).

It compares the gray value of each pixel to the gray values of its neighbours to generate an index code that represents the local pattern of each pixel. TheLBP code is implemented from the YT, XT, and XY planes of each pixel to augment information from the 3D spatial-temporal domain. The final LBP-TOP feature vector is formed by generating one histogram for each plane and then dividing it into a single histogram (Pfister *et al*., 2011).

LBP-TOP is one of the most widely used LBP approaches because Micro-Expression is time-dependent. As shown in the figure, a combination of LBP-TOP and a NN classifier is used to recognize Micro-Expressions. Because Micro-Expressions have a short duration, frame numbers in some Micro-Expression clips may be very limited. This creates a problem when it comes to parameters selection for LBP-TOP. To avoid this, a Temporal Interpolation Model (TIM) is used to temporally interpolate between frames, allowing the frame numbers to be varied (up-sampling and down-sampling) to promote the feature extraction process (Zhang and Arandjelović, 2021).

3) Deep Convolutional Neural Network (DCNN): The DCNN method was one of the first research-based facial expression analysis methods. A series of learning layers are created in this methodology to learn about each image's features. The input image will be passed through learning layers such as convolutional layers and completely connected layers (fully connected layers) for object detection, as shown in Figure 9. To extract features, each hidden layer in this work is a combination of convolution, Rectified Linear Units, Local Response Normalization, and pooling operations. The features extracted after the pooling process at the end of the first set of layers are used to identify facial expressions.

4) 3D Histograms of Oriented Gradients (3D HOG): Polikovsky *et al*. (Polikovsky, Kameda and Ohta, 2009) presented a scheme to spot facial Micro-Expressions. They manually annotated points on the face to divide them into 12 regions, then centered a rectangle on these points. The magnitudes of gradient projections in each of the three canonical directions are then used to build histograms across various regions, which are then used as features. The researchers conclude that each frame of the Micro-Expression image series contains only one Action Unit (AU) (Zhang and Arandjelović, 2021). To recognize motion in each region, 3D HOG was used. In most studies, 3D HOG features with weighted methods were combined with fuzzy classification to recognize Micro-Expressions.

5) Maximum Margin Projection with Tensor Representation (MMPTR): As third-order tensors, this algorithm can identify gait and Micro-Expressions. MMPTR may pursue a tensor-to-tensor projection that directly captures discriminative and geometry-preserving features from the original tensorial data by maximizing the inter-class Laplacian dispersion and suppressing the intra-class Laplacian scatter(Ben *et al*., 2016). MMPTR with Euclidean distance classifier provides high CASME dataset recognition accuracy.

6) Delaunay-Based Temporal Coding Mode (DTCM): In addition to the above-mentioned major feature extraction methods, DTCM is a rarely used method, but when combined with an SVM classifier, it has a higher prediction accuracy. Based on the feature points, Delaunay triangulation has been implemented. The face area is divided into several triangle-shaped sub-regions by this procedure. Normalization has been performed based on a standard face (neutral), which eliminates the irrelevance of personal appearance differences. Lu *et al*. has (Lu *et al*., 2015) coded the features space using Local Temporal Variations (LTVs), which computed the difference between a subregion's mean grayscale values

Copyright ©2021 belongs to Faculty of Technology, South Eastern University of Sri Lanka, University Park, Oluvil, #32360, Sri Lanka

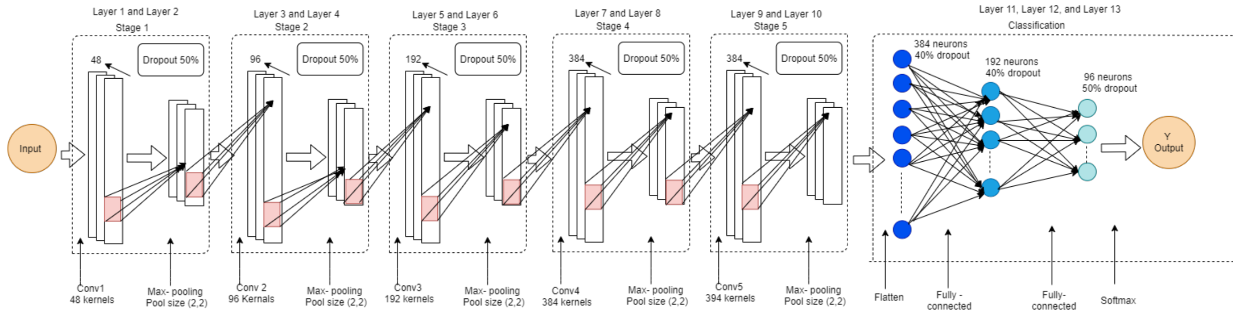

Figure 9: DCNN model to identify Micro-Expressions

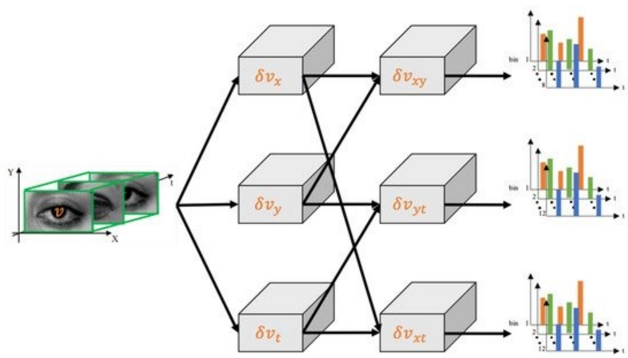

Figure 10: A conceptual overview of the 3DHOG descriptor extraction process for a facial cube (Zhang and Arandjelović, 2021)

and a subregion in a neighbour frame (Merghani *et al*., 2018).

B. Emotion Classification Techniques

Emotion classification is the last stage of this Micro-Emotion recognition system, in which the classifier categorizes the emotion classes such as happy, anger, sad, surprise, fear, disgust, and neutral.

1) Support Vector Machine -SVM: : Support Vector Machine is a prevalent classifier in the emotion recognition field, and it is also used in the Micro-Expression region field, along with other classifiers and extraction techniques. SVM can achieve near-optimal class separation. Using the features proposed, SVM is trained to perform facial expression classification. In general, SVM is the most general hyper-plane classification method, relying on Statistical Learning theory results to ensure high generalization performance. The SVM can be divided into two problems based on the number of classes from which the data samples are drawn: binary class problem (only two classes presented) and multi-class problem (Ex: Micro-expression systems have many classes).

To improve classification accuracy, SVM classifiers can be combined with other systems. For example, Wu *et al* (Li *et al*., 2013) have used the GentleSVM classifier, which is a combination of SVM and the Gentleboost algorithm, for detecting, and recognizing Micro-Expressions. However, that research method was not applied to real-life scenarios. Figure 11 shows how to use the previously discussed LBP-TOP feature extractor with SVM classifier to augment the micro-expression model.

2) K Nearest Neighbor (KNN): KNN algorithms process data and classify different data points using likeness measures (e.g. distance function). To identify its neighbors, a majority vote is used. SVM and K-Nearest Neighbor were used as classifiers in the CASME dataset to distinguish between 5 motions. During the classification procedure, the unlabeled query point is basically assigned the label of its K-Nearest Neighbors. The object is identified by a majority vote based on the labels of its K-Nearest Neighbors. Let K=1, the object is labeled as the closest object's class. When there is a binary classes problem, K must be an odd integer (only two distinct classes) (Zhang and Arandjelović, 2021). However, when performing multiclass classification, K may still be an odd integer. The most prominent distance function; Euclidean distance could be implemented for KNN, fter transforming each image to a fixed-length vector of real values (Zhang and Arandjelović, 2021). Consider Figure 12 below, it can distinguish three unique classes in this table by utilizing three different colors. When a new data 'x' arrives on the cartesian platform, it searches for its nearest values (if k = 3; looks at the closest 3 values) and chooses the class with the shortest distance to 'x'. Aside from the aforementioned mentioned above, numerous other technologies can be used in emotion recognition. Table III contains a summary of existing feature extraction and classification methods for various data types.

IV. Challenges and Limitations

There has been significant progress in the field of automated Micro-Expression recognition using machine learning in recent years. Several promising approaches have been proposed based on texture, gradient, and optical flow features. Conversely, numerous challenges and limitations have indeed been recognized in this research field. This section provides an outline of them.

1) The resolution of the image is critical for feature extraction. As a result, one of the most significant issues that might arise is the influence of Spatial-Temporal Settings in Data Collection.
2) There are only a few datasets that provide long video clips. Furthermore, not all datasets contain all universal emotional classes. Additionally, most of the time, not all object classes receive a balanced amount of datasets to

Copyright ©2021 belongs to Faculty of Technology, South Eastern University of Sri Lanka, University Park, Oluvil, #32360, Sri Lanka

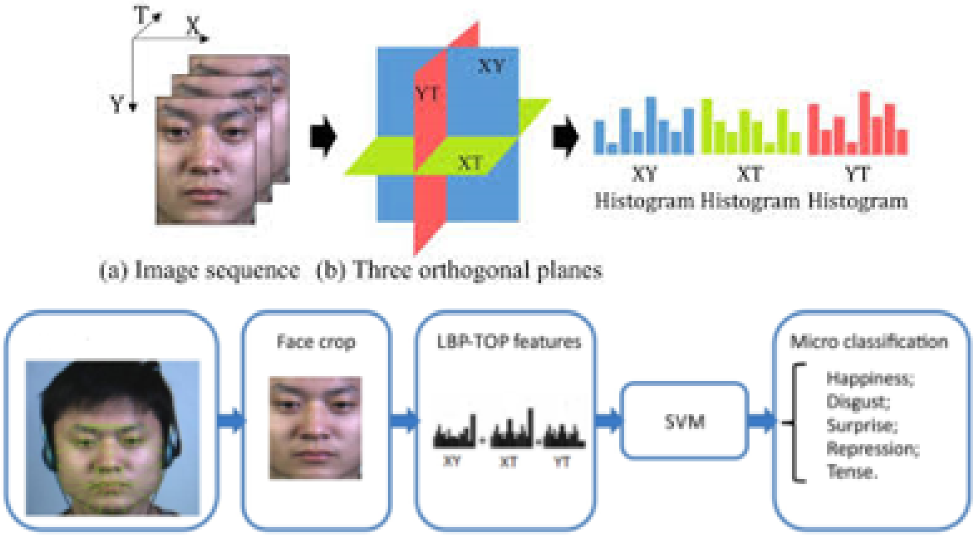

Figure 11: Micro-Expression classification using LBP-TOP and SVM (Yan *et al*., 2014)

Table III: Summary of different feature extraction and classification techniques

Dataset	Feature Extraction	Classifier	Results (Accuracy)
CASME	Tensor Independent Color Space (TICS) (Wang *et al*., 2014)	SVM	61.85%
	Main Directional Mean Optical-flow (MDMO) (Liu *et al*., 2016)	SVM	68.86%
	LBP-TOP with Histogram of Oriented Optical Flow (HOOF) (Zheng *et al*., 2016)	Relaxed K-SVD (RK-SVD)	69.04%
CASME II	TICS (Wang *et al*., 2014)	SVM	58.53%
	LBP-MOP (Mean Orthogonal Planes) (Wang *et al*., 2016)	SVM	66.80%
	Spatio-Temporal Local Binary Pattern with Integral Projection (STLBP-IP) (Huang *et al*., 2015)	SVM	59.51%
	MDMO (Liu *et al*., 2016)	SVM	67.37%
	LBP, HOG, and HIGO (Li *et al*., 2018)	LSVM	57.49%
	LBP-TOP and HOOF (Zheng *et al*., 2016)	RK-SVD	63.25%
	LBP-TOP (Wang *et al*., 2016)	SVM , KNN	75.30%
	Sparse Tensor Canonical Correlation Analysis (STCCA) (Wang *et al*., 2016)	KNN, SVM	Mean recognition accuracy: 38.39%

Table IV: Higher accurate feature extraction and classification methods on CASME dataset

Feature Extraction	Classifier	Results (Accuracy)
LBP-TOP	SVM and KNN	75.30%
3D HOG	Fuzzy	86.67%
MMPTR	Euclidean distance	80.20%
LBP-TOP+GDs (Gaussian derivatives) (Davison *et al*., 2015)	RF (Random Forests), SVM	92.6 % when RF used

Copyright ©2021 belongs to Faculty of Technology, South Eastern University of Sri Lanka, University Park, Oluvil, #32360, Sri Lanka

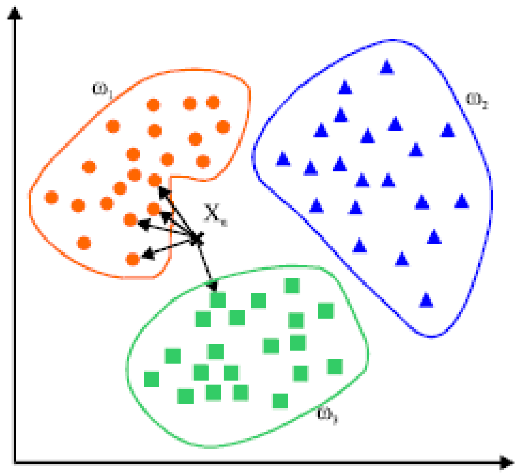

Figure 12: A 3-dimensional KNN classifier's feature space (Lee *et al.*, 2010)

train. Some emotional classes have a lot of data, while others don't.

3) It is difficult to distinguish between Micro-expressions and mark labels that correspond to a specific muscle movement pattern. In classification, attempts should made to eliminate the potential tendency of objective classes
4) The majority of accessible landmark detectors are not always accurate or consistent.
5) In real-time video emotion recognition, it is challenging to implement for every mind state.
6) Because Micro-Expressions are subtle, it is more likely that hidden features will be missed.
7) Significantly more memory is required, and both computation and testing are time-consuming.
8) A larger number of manually collected and labeled datasets are needed for the best accuracy.
9) It is challenging to identify when the same person has more different emotions at once.
10) Metric Standardization - Since the majority of datasets are unbalanced, it appears that stating the result in F-Measure (or F1-Score) would be the best option. Using the typical accuracy metric may result in a bias towards classes with a large number of samples, leading to an overestimation of the examined method's capability (Merghani *et al.*, 2018).
11) Micro-facial expressions arise when people tend disguise one's real feelings, and while knowing when someone is lying would be beneficial from an ethical standpoint, but it would take away your emotional freedom

V. Conclusion

The two most important phases in analyzing Micro-Expressions are feature extraction and classification. Choosing a detailed dataset should also be taken into account. This study has summarized a comprehensive review of datasets, feature extraction methods, and classification approaches for examining Micro-Expressions. In previous work, spontaneous datasets outperformed non-spontaneous datasets in terms of effectiveness. Furthermore, the vast majority of non-spontaneous datasets are no longer available for further analysis. As a result, this research has analyzed information concerning accessible spontaneous datasets. Many methods and combinations have been used in the field of Micro-Expression recognition out of many experimented methods over the past decades. However, not all methods are capable of providing greater accuracy in predicting micro-expressions. Therefore, if the CASME dataset was used, this research has concluded summarized feature extraction and classification combinations capable of achieving enough accuracy for the training model. This paper discusses the LBP, LBP-TOP, DCNN, 3DHOG, MMPTR, and DTCM filter methods for feature extraction which are identified as higher accurate techniques. Also, SVM, and KNN classification techniques can be provided as the best methods to classify extracted features. Table III summarizes some of the available techniques, and table IV tabulates high-accurate combinations. Micro-expression recognition is still in its early stages when compared to macro-expression recognition. Furthermore, implementing Micro-Expression recognition into practice in the real world is difficult. Micro-emotion recognition, on the other hand, has numerous real-world applications in security, health research, psychology, and other fields.

References

Ben, X., Zhang, P., Yan, R., Yang, M., Ge, G., 2016. Gait recognition and micro-expression recognition based on maximum margin projection with tensor representation. *Neural Computing and Applications 27*, 2629–2646. https://doi.org/10.1007/s00521-015-2031-8

C, G. (2020) "Attention Residual Network for Micro-expression Recognition Using Image Analysis", *Journal of Advanced Research in Dynamical and Control Systems, 12*(SP7), pp. 1261-1272. doi: 10.5373/jardcs/v12sp7/20202226.

Chen, Z. (2019) "Facial Expression Recognition Based on Local Features and Monogenic Binary Coding", *Informatica, 43*(1). doi: 10.31449/inf.v43i1.2716.

Ciolacu, M. (2014) "Theoretical Approaches Regarding the Psychophysiology of Emotions, Facial Expressions and Emotional Regulation", *Procedia - Social and Behavioral Sciences, 127*, pp. 748-752. doi: 10.1016/j.sbspro.2014.03.348.

Davison, A.K., Yap, M.H., Costen, N., Tan, K., Lansley, C., Leightley, D., 2015. Micro-Facial Movements: An Investigation on Spatio-Temporal Descriptors, in: Agapito, L., Bronstein, M.M., Rother, C. (Eds.), Computer Vision - ECCV 2014 Workshops, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 111–123. https://doi.org/10.1007/978-3-319-16181-5$_8$

Copyright ©2021 belongs to Faculty of Technology, South Eastern University of Sri Lanka, University Park, Oluvil, #32360, Sri Lanka

Davison, A. *et al*. (2018) "SAMM: A Spontaneous Micro-Facial Movement Dataset", *IEEE Transactions on Affective Computing, 9*(1), pp. 116-129. doi: 10.1109/taffc.2016.2573832.

Eulerian Video Magnification (2021). Available at: http://people.csail.mit.edu/mrub/evm/ (Accessed: 05 May 2021).

Guo, X. (2018) "Evaluation of Teaching Effectiveness based on Classroom Micro-Expression Recognition", *International Journal of Performability Engineering*. doi: 10.23940/ijpe.18.11.p33.28772885.

Huang, X. *et al*. (2015) "Facial Micro-Expression Recognition Using Spatiotemporal Local Binary Pattern with Integral Projection", 2015 IEEE International Conference on Computer Vision Workshop (ICCVW). doi: 10.1109/iccvw.2015.10.

Huang, X. *et al*. (2016) "Spontaneous facial micro-expression analysis using Spatiotemporal Completed Local Quantized Patterns", *Neurocomputing, 175*, pp. 564-578. doi: 10.1016/j.neucom.2015.10.096.

Pfister, T., Xiaobai Li, Zhao, G., Pietikainen, M., 2011. *Recognising spontaneous facial micro-expressions. IEEE*, pp. 1449–1456. https://doi.org/10.1109/ICCV.2011.6126401

Lee, L. *et al*. (2010) "A Review of Nearest Neighbor-Support Vector Machines Hybrid Classification Models", *Journal of Applied Sciences, 10*(17), pp. 1841-1858. doi: 10.3923/jas.2010.1841.1858.

Liong, S. *et al*. (2016) "Spontaneous subtle expression detection and recognition based on facial strain", Signal Liong, S. *et al*. (2016) "Spontaneous subtle expression detection and recognition based on facial strain", *Signal Processing: Image Communication, 47*, pp. 170-182.

Li, X., Pfister, T., Huang, X., Zhao, G., Pietikainen, M., 2013. A Spontaneous Micro-expression Database: Inducement, collection and baseline, in: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). Presented at the 2013 10th IEEE International Conference on Automatic Face Gesture Recognition (FG 2013), IEEE, Shanghai, China, pp. 1–6. https://doi.org/10.1109/FG.2013.6553717

Li, X. *et al*. (2018) "Towards Reading Hidden Emotions: A Comparative Study of Spontaneous Micro-Expression Spotting and Recognition Methods", *IEEE Transactions on Affective Computing, 9*(4), pp. 563-577. doi: 10.1109/taffc.2017.2667642.Shan, C., Gong, S., McOwan, P.W., 2009. Facial expression recognition based on Local Binary Patterns: A comprehensive study. Image and Vision Computing 27, 803–816. https://doi.org/10.1016/j.imavis.2008.08.005

Liu, Y.-J., Zhang, J.-K., Yan, W.-J., Wang, S.-J., Zhao, G., Fu, X., 2016. A Main Directional Mean Optical Flow Feature for Spontaneous Micro-Expression Recognition. *IEEE Transactions on Affective Computing 7*, 299–310. https://doi.org/10.1109/TAFFC.2015.2485205

Lu, Z., Luo, Z., Zheng, H., Chen, J., Li, W., 2015. A Delaunay-Based Temporal Coding Model for Micro-expression Recognition, in: Jawahar, C.V., Shan, S. (Eds.), Computer Vision - ACCV 2014 Workshops. Springer International Publishing, Cham, pp. 698–711. https://doi.org/10.1007/978-3-319-16631-5$_5$1

Merghani, W., Davison, A.K., Yap, M.H., 2018. *A Review on Facial Micro-Expressions Analysis: Datasets, Features and Metrics*. arXiv:1805.02397 [cs].

Papers with Code - Action Unit Detection (2021). Available at: https://paperswithcode.com/task/action-unit-detection/codeless (Accessed: 29 May 2021).

Peng, M. *et al*. (2019) "A Novel Apex-Time Network for Cross-Dataset Micro-Expression Recognition", 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). doi: 10.1109/acii.2019.8925525.

Pfister, T. *et al*. (2011) "Recognising spontaneous facial micro-expressions", 2011 International Conference on Computer Vision. doi: 10.1109/iccv.2011.6126401.

Polikovsky, S., Kameda, Y. and Ohta, Y. (2009) "Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor", 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009). doi: 10.1049/ic.2009.0244.

Shu, L. *et al*. (2018) "A Review of Emotion Recognition Using Physiological Signals", *Sensors, 18*(7), p. 2074. doi: 10.3390/s18072074.

Tian, Y., Kanade, T., Cohn, J.F., 2001. Recognizing Action Units for Facial Expression Analysis. *IEEE Trans Pattern Anal Mach Intell 23*, 97–115. https://doi.org/10.1109/34.908962

Wang, S.-J., Yan, W.-J., Li, X., Zhao, G., Fu, X., 2014. Micro-expression Recognition Using Dynamic Textures on Tensor Independent Color Space, in: 2014 22nd International Conference on Pattern Recognition. Presented at the 2014 22nd International Conference on Pattern Recognition, pp. 4678–4683. https://doi.org/10.1109/ICPR.2014.800

Wang, S.-J., Yan, W.-J., Sun, T., Zhao, G., Fu, X., 2016. Sparse tensor canonical correlation analysis for micro-expression recognition. *Neurocomputing 214*, 218–232. https://doi.org/10.1016/j.neucom.2016.05.083

Wang, Y. *et al*. (2015) "Efficient Spatio-Temporal Local Binary Patterns for Spontaneous Facial Micro-Expression Recognition", PLOS ONE, 10(5), p. e0124674. doi:

Copyright ©2021 belongs to Faculty of Technology, South Eastern University of Sri Lanka, University Park, Oluvil, #32360, Sri Lanka

10.1371/journal.pone.0124674.

Wang, Y. *et al*. (2016) "Effective recognition of facial micro-expressions with video motion magnification", *Multimedia Tools and Applications, 76*(20), pp. 21665-21690. doi: 10.1007/s11042-016-4079-6.

Wen-Jing Yan *et al*. (2013) "CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces", 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). doi: 10.1109/fg.2013.6553799.

Wu, Q., Shen, X., Fu, X., 2011. The Machine Knows What You Are Hiding: An Automatic Micro-expression Recognition System, in: D'Mello, S., Graesser, A., Schuller, B., Martin, J.-C. (Eds.), Affective Computing and Intelligent Interaction, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 152–162. https://doi.org/10.1007/978-3-642-24571-8$_1$6

Xiang, T., Leung, M. and Cho, S. (2008) "Expression recognition using fuzzy spatio-temporal modeling", *Pattern Recognition, 41*(1), pp. 204-216. doi: 10.1016/j.patcog.2007.04.021.

Yan, W. *et al*. (2014) "CASME II: An Improved Spontaneous Micro-Expression Database and the Baseline Evaluation", *PLoS ONE, 9*(1), p. e86041. doi: 10.1371/journal.pone.0086041.

Yao, L. *et al*. (2021) "Action Unit classification for facial expression recognition using active learning and SVM", *Multimedia Tools and Applications*. doi: 10.1007/s11042-021-10836-w.

Zhang, L., Arandjelović, O., 2021. Review of Automatic Microexpression Recognition in the Past Decade. *Machine Learning and Knowledge Extraction 3*, 414–434. https://doi.org/10.3390/make3020021

Zhao, Y., Xu, J., 2019. A Convolutional Neural Network for Compound Micro-Expression Recognition. *Sensors (Basel) 19*. https://doi.org/10.3390/s19245553

Zheng, H., Geng, X., Yang, Z., 2016. A Relaxed K-SVD Algorithm for Spontaneous Micro-Expression Recognition, in: Booth, R., Zhang, M.-L. (Eds.), PRICAI 2016: Trends in Artificial Intelligence, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 692–699. https://doi.org/10.1007/978-3-319-42911-3$_5$8

Zhu, X. *et al*. (2017) "Coupled source domain targetized with updating tag vectors for micro-expression recognition", *Multimedia Tools and Applications, 77*(3), pp. 3105-3124. doi: 10.1007/s11042-017-4943-z.

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. Te images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Copyright ©2021 belongs to Faculty of Technology, South Eastern University of Sri Lanka, University Park, Oluvil, #32360, Sri Lanka