

Article

Objective Classes for Micro-Facial Expression Recognition

Adrian K. Davison ¹, Walied Merghani ² and Moi Hoon Yap ^{3,*}

¹ Centre for Imaging Sciences, University of Manchester, Manchester M13 9PL, UK; adrian.davison@manchester.ac.uk

² Department of Computer Science, Sudan University of Science and Technology, Khartoum 11111, Sudan; waleed.a.mergani@gmail.com

³ School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, Manchester M15 6BH, UK

* Correspondence: M.Yap@mmu.ac.uk

Received: 1 September 2018; Accepted: 9 October 2018; Published: 15 October 2018



Abstract: Micro-expressions are brief spontaneous facial expressions that appear on a face when a person conceals an emotion, making them different to normal facial expressions in subtlety and duration. Currently, emotion classes within the CASME II dataset (Chinese Academy of Sciences Micro-expression II) are based on Action Units and self-reports, creating conflicts during machine learning training. We will show that classifying expressions using Action Units, instead of predicted emotion, removes the potential bias of human reporting. The proposed classes are tested using LBP-TOP (Local Binary Patterns from Three Orthogonal Planes), HOOF (Histograms of Oriented Optical Flow) and HOG 3D (3D Histogram of Oriented Gradient) feature descriptors. The experiments are evaluated on two benchmark FACS (Facial Action Coding System) coded datasets: CASME II and SAMM (A Spontaneous Micro-Facial Movement). The best result achieves 86.35% accuracy when classifying the proposed 5 classes on CASME II using HOG 3D, outperforming the result of the state-of-the-art 5-class emotional-based classification in CASME II. Results indicate that classification based on Action Units provides an objective method to improve micro-expression recognition.

Keywords: micro-facial expression; expression recognition; action unit

1. Introduction

A micro-facial expression is revealed when someone attempts to conceal their true emotion [1,2]. When they consciously realise that a facial expression is occurring, the person may try to suppress the facial expression because showing the emotion may not be appropriate [3]. Once the suppression has occurred, the person may mask over the original facial expression and cause a micro-facial expression. In a high-stakes environment, these expressions tend to become more likely as there is more risk to showing the emotion. The duration of a micro-expression is very short and is considered the main feature that distinguishes them from a facial expression [4], with the general standard being a duration of no more than 500 ms [5]. Other definitions of speed that have been studied show micro-expressions to last less than 200 ms, defined by Ekman and Friesen [6] as first to describe a micro-expression, 250 ms [7], less than 330 ms [8] and less than half a second [9].

Micro-facial expression analysis is less established and harder to implement due to being less distinct than normal facial expressions. Feature representations, such as Local Binary Patterns (LBP) [10–12], Histogram of Oriented Gradients (HOG) [13] and Histograms of Oriented Optical Flow (HOOF) [14], are commonly used to describe micro-expressions. Although micro-facial expression analysis is very difficult, the popularity in recent years has grown due to the potential applications

in security and interrogations [9,15–17], healthcare [18,19] and automatic detection in real-world applications where the detection accuracy of humans peaks around 40% [9].

Generally, the process of recognising normal facial expressions involves preprocessing, feature extraction and classification. Micro-expression recognition is not an exception, but the features extracted should be more descriptive due to the small movement in micro-expressions compared with normal expressions. One of the biggest problems faced by research in this area is the lack of publicly available datasets, on which the success in facial expression recognition [20] research largely relies. Gradually, datasets of spontaneously induced micro-expression have been developed [21–24], but earlier research was centred around posed datasets [25,26].

Eliciting spontaneous micro-expression is a real challenge because it can be very difficult to induce the emotions in participants and also get them to conceal them effectively in a lab-controlled environment. Micro-expression datasets need decent ground truth labelling with Action Units (AUs) using the Facial Action Coding System (FACS) [27]. FACS objectively assigns AUs to the muscle movements of the face. If any classification of movements takes place for micro-facial expressions, it should be done with AUs and not only emotions. Emotion classification requires the context of the situation for an interpreter to make a meaningful interpretation. Most spontaneous micro-expression datasets have FACS ground truth labels and estimated or predicted emotion. These have been annotated by an expert and self-reports written by participants.

We contend that using AUs to classify micro-expressions gives more accurate results than using predicted emotion categories. By organising the AUs of the two most recent FACS coded state-of-the-art datasets, CASME II [23] and SAMM [24], into objective classes, we ensure that the learning methods train on specific muscle movement patterns and therefore increase accuracy. Yan et al. [28] also state that it is inappropriate to categorise micro-expressions into emotion categories, and that using FACS AU research to inform the eventual emotional classification. To date, experiments on micro-expression recognition using categories based purely on AU movements, has not been completed. Additionally, the SAMM dataset was designed for micro-movement analysis rather than recognition. We contribute by completing recognition experiments on the SAMM dataset for the first time with three features previously used for micro-expression analysis: LBP-TOP [12], HOOF [14] and HOG 3D [13,25]. Further, the proposed objective classes could inform future research on the importance of objectifying movements of the face.

The remainder of this paper is divided into the following sections; Section 2 discusses the background of two of the FACS coded state-of-the-art datasets developed for micro-expression analysis and the related work in micro-expression recognition; Section 3 describes the methodology; Section 4 presents the results and discusses the effects of applying objective classification to a micro-expression recognition task; Section 5 concludes this paper and discusses future work.

2. Background

This section will describe two datasets which are used in the experiments for this paper. A comparative summary of the datasets can be seen in Table 1. Previously developed micro-expression recognition systems are also discussed using established features to represent each micro-expression.

Table 1. A summary of the different features of the CASME II and SAMM datasets.

Feature	CASME II [23]	SAMM [24]
Micro-Movements	247 *	159
Participants	35	32
Resolution	640 × 480	2040 × 1088
Facial Resolution	280 × 340	400 × 400
FPS	200	200
Spontaneous/Posed	Spontaneous	Spontaneous
FACS Coded	Yes	Yes
No. Coders	2	3
Emotion Classes	5	7
Mean Age (SD)	22.03 (SD = 1.60)	33.24 (SD = 11.32)
Ethnicities	1	13

* This is the original amount of movements used in [23], however we use a larger set of 255 provided by the dataset.

2.1. CASME II

CASME II was developed by Yan et al. [23] and refers to the Chinese Academy of Sciences Micro-expression Database II, which was preceded by CASME [22] with major improvements. All samples in CASME II are spontaneous and dynamic micro-expressions with a high frame rate (200 fps). There are a few frames kept before and after each micro-expression to make it suitable for detection experiments. The resolution of samples is 640 × 480 pixels for recording which saved as MJPEG and resolution about 280 × 340 pixels for cropped facial area. The participants' facial expressions were elicited in a well-controlled laboratory environment. The dataset contains 255 micro-expressions (gathered from 35 participants) and were selected from nearly 3000 facial movements and have been labelled with AUs based on FACS. Only 247 movements were used in the original experiments on CASME II [23]. The inter-coder reliability of the FACS codes within the dataset is 0.846. Flickering light was avoided in the recordings and highlights to regions of the face were reduced. However, there were some limitations: Firstly, the materials used for eliciting micro-expression are video episodes which can have different meanings to different people, for example eating worms may not always disgust someone. Secondly, micro-expressions are elicited under one specific lab situation. There was some difficulty in eliciting some types of facial expressions in laboratory situations, such as sadness.

When analysing the FACS codes of the CASME II dataset, it was found that there are many conflicts to the coded AUs and the estimated emotions. These inconsistencies do not help when attempting to train distinct machine learning classes, and adds further justification for the proposed introduction of new classes based on AUs only.

For example, Subject 11 with the micro-expression clip filename of 'EP19_03f', was coded as an AU4 in the 'others' estimated emotion category (shown in Figure 1). However, Subject 26 with the micro-expression clip filename of 'EP18_50', was also coded with AU4 but in the 'disgust' estimated emotion category (shown in Figure 2). As can be seen in the apex frame (centre image) of both Figures 1 and 2, AU4, the lowering of the brow, is present. Having the same movement in different categories is likely to have an effect on any training stage of machine learning.



Figure 1. Sample frames showing Subject 11’s micro-expression clip ‘EP19_03f’ that was coded as an AU4 in the ‘others’ category (©Xiaolan Fu).



Figure 2. Sample frames showing Subject 26’s micro-expression clip ‘EP18_50’ that was coded as an AU4 in the ‘disgust’ category (©Xiaolan Fu).

2.2. SAMM

The Spontaneous Actions and Micro-Movements (SAMM) [24] dataset is the first high-resolution dataset of 159 micro-movements induced spontaneously with the largest variability in demographics. To obtain a wide variety of emotional responses, the dataset was created to be as diverse as possible. A total of 32 participants were recruited for the experiment with a mean age of 33.24 years (SD: 11.32, ages between 19 and 57), and an even gender split of 16 male and female participants. The inter-coder reliability of the FACS codes within the dataset is 0.82, and was calculated by using a slightly modified version of the inter-reliability formula, found in the FACS Investigator’s Guide [29], to account for three coders rather than two.

The inducement procedure was based on the seven basic emotions [1] and recorded at 200 fps. As part of the experimental design, each video stimuli was tailored to each participant, rather than obtaining self reports during or after the experiment. This allowed for particular videos to be chosen and shown to participants for optimal inducement potential. The experiment comprised of 7 stimuli used to induce emotion in the participants who were told to suppress their emotions so that micro-facial movements might occur. To increase the chance of this happening, a prize of £50 was offered to the participant that could hide their emotion the best, therefore introducing a high-stakes situation [1,2]. Each participant completed a questionnaire prior to the experiment so that the stimuli could be tailored to each individual to increase the chances of emotional arousal.

The SAMM dataset was originally designed to investigate micro-facial movements by analysing muscle movements of the face rather than recognising distinct classes [30]. We are the first to categorise SAMM based on the FACS AUs and then use these categories for micro-facial expression recognition.

2.3. Related Work

Currently, there are three features which many micro-expression recognition approaches rely on: Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG) and Histogram of Oriented Optical Flow (HOOOF) based. We will discuss different methods that use these features in recent work on micro-expression recognition. Further, specific important micro-expression research is discussed.

As an extension to the original Local Binary Pattern (LBP) [11] operator, Local Binary Patterns on Three Orthogonal Planes (LBP-TOP) was proposed by Zhao et al. [12] demonstrated to be effective for dynamic texture and facial expression analysis in the spatial–temporal domain. Given a video sequence of time length T , usually it can be thought as a stack of XY planes along the time axis T ,

but also it can be thought as three planes XY, XT and YT. These provide information about space and time transition. The basic idea of LBP-TOP is similar to LBP, the difference being that LBP-TOP extracts features from all three planes which will be combined into a single feature vector.

Yan et al. [23] carried out the first micro-expression recognition experiment on the CASME II dataset. LBP-TOP [12] was used to extract the features and Support Vector Machine (SVM) [31] was employed as the classifier. The radii varied from 1 to 4 for X and Y, and from 2 to 4 for T (T = 1 was not considered due to little change between two neighbouring frames at 200 fps), with classification occurring between five main categories of emotions provided in this experiment (happiness, disgust, surprise, repression and others).

Davison et al. [32] used the LBP-TOP feature to differentiate between movements and neutral sequences, attempting to avoid bias when classifying with an SVM.

The performance of [23] on recognising micro-expressions in 5-classes with LBP-TOP features extraction, achieved a best result of 63.41% accuracy, using leave-one-out cross-validation. This result is an average for recent micro-expression recognition research, and is likely due to the way micro-expressions are categorised. Of the 5-class in the CASME II dataset, 102 were classed as 'others', which denoted movements not suited for the other categories but related to emotion. The next highest category was 'disgust' with 60 movements, showing that the 'others' class made the categorisation imbalanced. Further, the categorisation was not based solely on AUs due to micro-expressions being short in duration and low in intensity, but also based on the participant's self-reporting. By classifying micro-expressions in this way, features are unlikely to exhibit a pattern, and therefore perform poorly during the recognition stage as can be seen in the other performance results. For example, in [23], the highest results is 63.41%, which is still relatively low.

More recently, LBP-TOP was used as a base feature for micro-expression recognition with integral projection [33,34]. These representations attempt to improve discrimination between micro-expression classes and therefore improve recognition rates. Polikovskiy et al. [25] used a 3D gradient histogram descriptor (HOG 3D) to recognise posed micro-facial expressions from high-speed videos. The paper used manually marked up areas that are relevant to FACS-based movement so that unnecessary parts of the face are left out. This does mean that the method of classifying movement in these subjectively selected areas is time-consuming and would not suit a real-time application like interrogation. The spatio-temporal domain is explored highlighting the importance of the temporal plane in micro-expressions, however the bin selection for the XY plane is 8 and the XT, YT planes have been set to 12. The number of bins selected represents the different directions of movement in each plane.

For HOOOF-based methods, a Main Direction Mean Optical Flow (MDMO) feature was proposed by Liu et al. [35] for micro-facial expression recognition using SVM as a classifier. The method of detection also uses 36 regions, partitioned using 66 facial points on the face, to isolate local areas for analysis, but keeping the feature vector small for computational efficiency. The best result on the CASME II dataset was 67.37% using leave-one-subject-out cross validation.

The basic HOOOF descriptor was also used by Li et al. [36] as a comparative feature when spotting micro-expressions and then performing recognition. This is the first automatic micro-expression system which can spot and recognise micro-expressions from spontaneous video data, and be comparable to human performance.

Using Robust Principal Component Analysis (RPCA) [37], Wang et al. [38] extract the sparse information from micro-expression data, and then use Local Spatiotemporal Directional Features, based on LBP-TOPs dynamic features, to extract the subtle motion on the face from 16 ROIs of local importance to facial expression motion.

A novel colour space model was created named Tensor Independent Color Space (TICS), that helps recognise micro-expressions [39]. By extracting the LBP-TOP features of independent colour components, micro-expression clips can be better recognised than the RGB space.

Huang et al. [40] proposed Spatio-Temporal Completed Local Quantization Patterns (STCLQP), which extracts the sign, magnitude and orientation of the micro-expression data, then an efficient vector

quantization and codebook selection are developed in both the appearance and temporal domains for generalising classical pattern types. Finally, using the developed codebooks, spatio-temporal features of sign, magnitude and orientation components are extracted and fused, with experiments being run on SMIC, CASME and CASME II.

By exploiting the sparsity in the spatial and temporal domains of micro-expressions, a Sparse Tensor Canonical Correlation Analysis was proposed for micro-expression characteristics [41]. This method reduces the dimensionality of micro-expression data and enhances LBP coding to find a subspace to maximise the correlation between micro-expression data and their corresponding LBP code.

Liong et al. [42] investigate the use of only two frames from a micro-expression clip: the onset and the apex frame. By only using a couple of frames, a good accuracy is achieved when using the proposed Bi-Weighted Oriented Optical Flow feature to encode the expressiveness of the apex frame.

As micro-movements on the face are heavily affected by the global movements of a person’s head, Xu et al. [43] propose a Facial Dynamics Map to distinguish between what is a micro-expression and what would be classed as a non-micro-expression. The facial surface movement between adjacent frames is predicted using optical flow. The movements are then extracted in a coarse-to-fine manner, indicating different levels of facial dynamics. This step is used to differentiate micro-movements from anything else. Finally, an SVM is used for both identification and categorisation.

Wang et al. [44] recently proposed a Main Directional Maximal Difference (MDMD) method that uses the magnitude maximal difference in the main direction of optical flow features to find when facial movements occur. These movements can be used for both micro-expressions and macro-expressions to find the onset, apex and offset of a movement within the context of each examined clip.

3. Methodology

To overcome the conflicting classes in CASME II, we restructure the classes around the AUs that have been FACS coded. Using EMFACS [29], a list of AUs and combinations are proposed for a fair categorisation of the SAMM [24] and CASME II [23] datasets. Categorising in this way removes the bias of human reporting and relies on the ground truth movement data, feature representation and recognition technique for each micro-expression clip. Table 2 shows 7 classes and the corresponding AUs that have been assigned to that class. Classes I–VI are linked with happiness, surprise, anger, disgust, sadness and fear. Class VII relates to contempt and other AUs that have no emotional link in EMFACS [29]. It should be noted that the classes do not directly correlate to being these emotions, however the links used are informed from previous research [27,29,45].

Table 2. Each class represents AUs that can be linked to emotion.

Class	Action Units
I	AU6, AU12, AU6+AU12, AU6+AU7+AU12, AU7+AU12
II	AU1+AU2, AU5, AU25, AU1+AU2+AU25, AU25+AU26, AU5+AU24
III	A23, AU4, AU4+AU7, AU4+AU5, AU4+AU5+AU7, AU17+AU24, AU4+AU6+AU7, AU4+AU38
IV	AU10, AU9, AU4+AU9, AU4+AU40, AU4+AU5+AU40, AU4+AU7+AU9, AU4 +AU9+AU17, AU4+AU7+AU10, AU4+AU5+AU7+AU9, AU7+AU10
V	AU1, AU15, AU1+AU4, AU6+AU15, AU15+AU17
VI	AU1+AU2+AU4, AU20
VII	Others

Each movement in both datasets were classified based on the AU categories of Table 2, with the resulting frequency of movements being shown in Table 3.

Table 3. The total number of movements assigned to the new classes for both SAMM and CASME II.

Class	CASME II	SAMM	Total
I	25	24	49
II	15	13	28
III	99	20	119
IV	26	8	34
V	20	3	23
VI	1	7	8
VII	69	84	153
Total	255	159	415

Micro-expression recognition experiments are run on two datasets: CASME II and SAMM. For this experiment, three types of feature representations are extracted from a sequence of grey images which represent the micro-expressions. These image sequences are divided into 5×5 blocks that are non-overlapping. The LBP-TOP features [12] radii parameters for X, Y and T are set to 1, 1 and 4 respectively and all neighbours in three planes set to 4. The HOG 3D [25] and HOOOF [14] features are set to the parameters described in the original implementations.

Sequential Minimal Optimization (SMO) [46] is used in the classification phase with ten-fold cross validation and leave-one-subject-out (LOSO) to classify between I–V, I–VI and I–VII classes. SMO is a fast algorithm for training SVMs, and provide a solution to solving very large quadratic programming (QP) problems, which are required to train SVMs. SMO avoid time-consuming QP calculations by breaking them down into smaller pieces. Doing this allows for the classification task to be completed much faster than using traditional SVMs [46].

4. Results

Evidence to support the proposed AU-based categories can be seen in the confusion matrix in Figure 3. A high proportion of micro-expressions have been classified as ‘others’, for example 28.95% of the ‘happiness’ and 28.57% of the ‘disgust’ categories are classified as ‘others’ respectively. The original chosen emotions, including many placed in the ‘others’ category, leads to a lot of conflict at the recognition stage. It should be noted that the CASME II dataset [23] included self-reporting, which adds another layer of complexity during classification.

The proposed classes I–V classification results using LBP-TOP can be seen in the confusion matrix in Figure 4. In contrast, the classification rates are more stable and outperforming the original classes overall. The results are by no means perfect, however it shows that the most logical direction is to use objective classes based on AUs rather than estimated emotion categories. Further investigation using an objective selection of FACS-based regions [47] supports this with AUC results for detecting relevant movements to be 0.7512 and 0.7261 on SAMM and CASME II, respectively.

	Happiness	Disgust	Surprise	Repression	Others
Happiness	39.47	4.76	0.00	30.77	5.77
Disgust	2.63	60.32	12.50	7.69	18.27
Surprise	5.26	4.76	75.00	0.00	1.92
Repression	23.68	1.59	4.17	50.00	2.88
Others	28.95	28.57	8.33	11.54	71.15

Figure 3. Confusion matrix of the original CASME II classes using the LBP-TOP feature, using SMO as a classifier.

	I	II	III	IV	V
I	76.19	10.53	2.94	11.11	6.25
II	0.00	42.11	1.96	7.41	18.75
III	4.76	21.05	83.33	29.63	6.25
IV	4.76	15.79	8.82	48.15	0.00
V	14.29	10.53	2.94	3.70	68.75

Figure 4. Confusion matrix of the proposed classes I–V on the CASME II dataset using the LBP-TOP feature and SMO as a classifier.

Table 4 shows the experimental results on CASME II with each result metric being a weighted average calculation to account for imbalanced numbers within classes. Each experiment was completed for each feature and within the original classes defined in [23] and the proposed classes. To compare with the state-of-the-art 5-class emotional-based classification in CASME II testing have done to classify 5 proposed classes(I–V). In addition to that and for more information and details and because we have 7 proposed classes testing to classify 6 classes(I–VI) and 7 classes(I–VII) also has been done and reporting as shown. Both the ten-fold cross-validation results and leave-one-subject-out (LOSO) are shown.

The top performing feature achieves a weighted accuracy score of 86.35% for the HOG 3D feature in the proposed classes I–V. This shows a large improvement over the original classes which achieved 80.93% for the same feature. Using LOSO, the results were comparable with the original classes. The highest accuracy was 76.60% from the HOOOF feature, in the proposed I–VII classes. For the CASME II dataset results, using LBP-TOP and ten-fold cross-validation, the original method outperformed

the classes I–VI and I–VII. In addition, for HOG3D LOSO, the original method outperforms in classes I–VII when using F-measure as a measurement.

The experiment based on the same conditions were then repeated for SAMM and can be seen in Table 5. Overall the recognition rates were good for SAMM, with the best result achieving an accuracy of 81.93% using LBP-TOP in I–VI classes for ten-fold cross validation. The best result using LOSO was from the HOG 3D feature, in the proposed I–VII classes and achieved 63.93%, however due to the lower amount of micro-expressions within the SAMM dataset compared with CASME II, the LOSO results were lower.

Some results show that using LOSO, HOOF outperforms in CASME II while HOG3D outperforms in SAMM and in CASME II using LOSO, the HOOF feature achieves a higher accuracy for classes I–VII over I–VI, but not for the F-measure metric. Explanations of this comes down to the data, and how large some variations of the settings, such as resolution and capture methods, are set. The imbalance of data, specifically the low amounts of micro-expression data, can skew LOSO results with low amounts of testing and training. This shows how using LOSO for micro-expression recognition is difficult to quantify with a fair amount of significance. Further data collection of spontaneous micro-expressions is required to rectify this.

Table 4. Results on the CASME II dataset showing each feature, proposed classes, and the original classes defined in [23] for comparison.

Feature	Class	Ten-fold Cross-Validation					Leave-One-Subject-Out (LOSO)				
		Accuracy (%)	TPR	FPR	F-Measure	AUC	Accuracy (%)	TPR	FPR	F-Measure	AUC
LBP-TOP	Original	77.17	0.56	0.22	0.53	0.74	68.24	0.49	0.17	0.48	0.63
	I–V	77.94	0.63	0.33	0.58	0.70	67.80	0.54	0.14	0.51	0.44
	I–VI	76.84	0.59	0.32	0.55	0.69	67.94	0.53	0.14	0.51	0.44
	I–VII	76.13	0.50	0.23	0.45	0.70	61.92	0.39	0.17	0.35	0.63
HOOF	Original	78.83	0.61	0.19	0.60	0.78	68.36	0.51	0.24	0.49	0.61
	I–V	82.70	0.69	0.22	0.67	0.80	69.64	0.59	0.18	0.56	0.47
	I–VI	82.41	0.68	0.23	0.66	0.79	73.52	0.62	0.18	0.60	0.47
	I–VII	83.94	0.64	0.14	0.63	0.79	76.60	0.57	0.14	0.55	0.72
HOG3D	Original	80.93	0.62	0.14	0.62	0.79	59.59	0.38	0.24	0.35	0.50
	I–V	86.35	0.72	0.13	0.72	0.84	69.53	0.56	0.18	0.51	0.40
	I–VI	83.49	0.68	0.16	0.67	0.80	69.87	0.56	0.18	0.51	0.40
	I–VII	82.59	0.58	0.12	0.58	0.79	61.33	0.39	0.30	0.31	0.51

Table 5. Results on the SAMM dataset showing each feature and proposed classes.

Feature	Class	Ten-fold Cross-Validation					Leave-One-Subject-Out (LOSO)				
		Accuracy (%)	TPR	FPR	F-Measure	AUC	Accuracy (%)	TPR	FPR	F-Measure	AUC
LBP-TOP	I–V	79.21	0.54	0.16	0.51	0.74	44.70	0.38	0.19	0.35	0.31
	I–VI	81.93	0.55	0.13	0.52	0.74	45.89	0.34	0.17	0.31	0.36
	I–VII	79.52	0.57	0.18	0.56	0.74	54.93	0.42	0.22	0.39	0.40
HOOF	I–V	78.95	0.56	0.16	0.55	0.74	42.17	0.32	0.06	0.33	0.32
	I–VI	79.53	0.52	0.15	0.51	0.73	40.89	0.28	0.07	0.27	0.35
	I–VII	72.80	0.52	0.32	0.50	0.65	60.06	0.49	0.25	0.48	0.30
HOG3D	I–V	77.18	0.51	0.17	0.49	0.74	34.16	0.22	0.15	0.22	0.24
	I–VI	79.41	0.48	0.15	0.45	0.71	36.39	0.19	0.14	0.19	0.26
	I–VII	79.09	0.59	0.25	0.55	0.71	63.93	0.50	0.22	0.44	0.30

5. Limitations

We agree that while the reliable AU indicators used within the EMFACS system are not perfectly able to give an objective stance, emotion mapping is limited to these forms of typical classification. Individual AU classification could be an interesting way to be purely objective on an individual AU

level. A disadvantage to this would be a separation from how emotion can be defined and how a single AU may not be able to describe what a person may (or may not) be feeling. Further, individual AU example movements from datasets are limited in that some are much more common (e.g., AU4) than others (AU1).

6. Conclusions

We show that restructuring micro-expression classes objectively around the AUs recognition results outperforms the state-of-the-art, emotion-based classification approaches. As micro-expressions are so subtle, the way to categorise is as objectively as possible, so using AU codes is the most logical. Categorising using a combination of AUs and self-reports [23] can cause many conflicts when training a machine learning method. Further, dataset imbalances can be very detrimental to machine learning algorithms, and this is further emphasised with the relatively low amount of movements in both datasets. Future work will probably look into the effect of using more modern features, with AUs classification to improve on the recognition accuracy. This could include the MDMO feature [35], local wrinkle feature [48] and the feature extraction methods described by Wang et al. [49].

Further work can be done to improve micro-facial expression datasets. Firstly, more datasets or expanding previous sets would be a simple improvement that can help move the research forward faster. Secondly, a standard procedure on how to maximise the amount of micro-movements induced spontaneously in laboratory controlled experiments would be beneficial. If collaboration between established datasets and researchers from psychology occurred, dataset creation would be more consistent.

Deep learning has emerged as a new area of machine learning research [50–52], and micro-expression analysis has yet to exploit this trend. Unfortunately, the amount of high-quality spontaneous micro-expression data is low and deep learning requires a large amount of data to work well [51]. Many video-based datasets previously used have over 10,000 video samples [53] and even over 1 million actions extracted from YouTube videos [54]. A real effort to gather spontaneous micro-expression data is required for deep learning approaches to be effective in the future.

Author Contributions: A.K.D. carried out the design of the study, the re-classification of the Action Units grouping and drafted the manuscript (The tasks was completed when A.K.D. was in Manchester Metropolitan University). W.M. conducted the experiments, analysed the data and drafted the manuscript. M.H.Y. designed the study, developed the theory, assisted development and testing and edited the manuscript. All the authors have read and approved this version of the manuscript. Conceptualization, A.K.D. and M.H.Y.; Data curation, W.M.; Funding acquisition, M.H.Y.; Methodology, A.K.D.; Software, W.M.; Supervision, M.H.Y.; Writing—original draft, A.K.D.; Writing—review & editing, W.M. and M.H.Y.

Funding: This work was completed in Manchester Metropolitan University on a “Future Research Leaders Programme” awarded to M.H.Yap. M.H.Yap is a Royal Society Industry Fellow.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ekman, P. *Emotions Revealed: Understanding Faces and Feelings*; Phoenix: Nairobi, Kenya, 2004.
2. Ekman, P. Lie Catching and Microexpressions. In *The Philosophy of Deception*; Martin, C.W., Ed.; Oxford University Press: New York, NY, USA, 2009; pp. 118–133.
3. Matsumoto, D.; Yoo, S.H.; Nakagawa, S. Culture, emotion regulation, and adjustment. *J. Pers. Soc. Psychol.* **2008**, *94*, 925. [[CrossRef](#)] [[PubMed](#)]
4. Shen, X.B.; Wu, Q.; Fu, X.L. Effects of the duration of expressions on the recognition of microexpressions. *J. Zhejiang Univ. Sci. B* **2012**, *13*, 221–230. [[CrossRef](#)] [[PubMed](#)]
5. Yan, W.J.; Wu, Q.; Liang, J.; Chen, Y.H.; Fu, X. How Fast are the Leaked Facial Expressions: The Duration of Micro-Expressions. *J. Nonverbal Behav.* **2013**, *37*, 217–230. [[CrossRef](#)]
6. Ekman, P.; Friesen, W.V. Nonverbal leakage and clues to deception. *Psychiatry* **1969**, *32*, 88–106. [[CrossRef](#)] [[PubMed](#)]
7. Ekman, P. *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*; Norton: New York, NY, USA, 2001.

8. Ekman, P.; Rosenberg, E.L. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*; Series in Affective Science; Oxford University Press: New York, NY, USA, 2005.
9. Frank, M.G.; Maccario, C.J.; Govindaraju, V.I. Behavior and Security. In *Protecting Airline Passengers in the Age of Terrorism*; Greenwood Pub. Group: Santa Barbara, CA, USA, 2009.
10. Ojala, T.; Pietikäinen, M.; Harwood, D. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognit.* **1996**, *29*, 51–59. [[CrossRef](#)]
11. Ojala, T.; Pietikäinen, M.; Mäenpää, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987. [[CrossRef](#)]
12. Zhao, G.; Pietikäinen, M. Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 915–928. [[CrossRef](#)] [[PubMed](#)]
13. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.
14. Chaudhry, R.; Ravichandran, A.; Hager, G.; Vidal, R. Histograms of oriented optical flow and Binet-Cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009), Miami, FL, USA, 20–25 June 2009; pp. 1932–1939. [[CrossRef](#)]
15. O’Sullivan, M.; Frank, M.G.; Hurley, C.M.; Tiwana, J. Police lie detection accuracy: The effect of lie scenario. *Law Hum. Behav.* **2009**, *33*, 530. [[CrossRef](#)] [[PubMed](#)]
16. Frank, M.; Herbasz, M.; Sinuk, K.; Keller, A.M.; Kurylo, A.; Nolan, C. *I See How You Feel: Training Laypeople and Professionals to Recognize Fleeting Emotions*; International Communication Association: New York, NY, USA, 2009.
17. Yap, M.H.; Ugail, H.; Zwigelaar, R. A database for facial behavioural analysis. In Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, 22–26 April 2013; pp. 1–6.
18. Hopf, H.C.; Muller-Forell, W.; Hopf, N.J. Localization of emotional and volitional facial paresis. *Neurology* **1992**, *42*, 1918–1918. [[CrossRef](#)] [[PubMed](#)]
19. Cohn, J.F.; Kruez, T.S.; Matthews, I.; Yang, Y.; Nguyen, M.H.; Padilla, M.T.; Zhou, F.; De La Torre, F. Detecting depression from facial actions and vocal prosody. In Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops (ACII 2009), Amsterdam, The Netherlands, 10–12 September 2009; pp. 1–7. [[CrossRef](#)]
20. Yap, M.H.; Ugail, H.; Zwigelaar, R. Facial Behavioral Analysis: A Case Study in Deception Detection. *Br. J. Appl. Sci. Technol.* **2014**, *4*, 1485. [[CrossRef](#)]
21. Li, X.; Pfister, T.; Huang, X.; Zhao, G.; Pietikäinen, M. A Spontaneous Micro-expression Database: Inducement, Collection and Baseline. In Proceedings of the 10th IEEE International Conference on Automatic Face and Gesture Recognition, Shanghai, China, 22–26 April 2013.
22. Yan, W.J.; Wu, Q.; Liu, Y.J.; Wang, S.J.; Fu, X. CASME Database: A dataset of spontaneous micro-expressions collected from neutralized faces. In Proceedings of the IEEE Conference on Automatic Face and Gesture Recognition, Shanghai, China, 22–26 April 2013.
23. Yan, W.J.; Li, X.; Wang, S.J.; Zhao, G.; Liu, Y.J.; Chen, Y.H.; Fu, X. CASME II: An Improved Spontaneous Micro-Expression Database and the Baseline Evaluation. *PLoS ONE* **2014**, *9*, e86041. [[CrossRef](#)] [[PubMed](#)]
24. Davison, A.K.; Lansley, C.; Costen, N.; Tan, K.; Yap, M.H. SAMM: A Spontaneous Micro-Facial Movement Dataset. *IEEE Trans. Affect. Comput.* **2018**, *9*, 116–129. [[CrossRef](#)]
25. Polikovskiy, S.; Kameda, Y.; Ohta, Y. Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor. In Proceedings of the 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009), London, UK, 3 December 2009; pp. 16–21.
26. Shreve, M.; Godavarthy, S.; Goldgof, D.; Sarkar, S. Macro- and micro-expression spotting in long videos using spatio-temporal strain. In Proceedings of the 2011 IEEE International Conference on Automatic Face Gesture Recognition and Workshops (FG 2011), Santa Barbara, CA, USA, 21–25 March 2011; pp. 51–56. [[CrossRef](#)]
27. Ekman, P.; Friesen, W.V. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*; Consulting Psychologists Press: Palo Alto, CA, USA, 1978.

28. Yan, W.J.; Wang, S.J.; Liu, Y.J.; Wu, Q.; Fu, X. For micro-expression recognition: Database and suggestions. *Neurocomputing* **2014**, *136*, 82–87. [[CrossRef](#)]
29. Ekman, P.; Friesen, W.V. *Facial Action Coding System: Investigator's Guide*; Consulting Psychologists Press: Palo Alto, CA, USA, 1978.
30. Davison, A.K.; Yap, M.H.; Lansley, C. Micro-Facial Movement Detection Using Individualised Baselines and Histogram-Based Descriptors. In Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Kowloon, China, 9–12 October 2015; pp. 1864–1869. [[CrossRef](#)]
31. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. doi:10.2262/7411411. [[CrossRef](#)]
32. Davison, A.K.; Yap, M.H.; Costen, N.; Tan, K.; Lansley, C.; Leightley, D. Micro-facial Movements: An Investigation on Spatio-Temporal Descriptors. In *13th European Conference on Computer Vision (ECCV)*; Springer: Cham, Switzerland, 2014.
33. Huang, X.; Wang, S.J.; Zhao, G.; Piteikainen, M. Facial Micro-Expression Recognition Using Spatiotemporal Local Binary Pattern With Integral Projection. In Proceedings of the The IEEE International Conference on Computer Vision (ICCV) Workshops, Santiago, Chile, 7–13 December 2015.
34. Huang, X.; Wang, S.; Liu, X.; Zhao, G.; Feng, X.; Pietikainen, M. Spontaneous Facial Micro-Expression Recognition using Discriminative Spatiotemporal Local Binary Pattern with an Improved Integral Projection. *arXiv* **2016**, arXiv:1608.02255.
35. Liu, Y.J.; Zhang, J.K.; Yan, W.J.; Wang, S.J.; Zhao, G.; Fu, X. A Main Directional Mean Optical Flow Feature for Spontaneous Micro-Expression Recognition. *IEEE Trans. Affect. Comput.* **2016**, *7*, 299–310. [[CrossRef](#)]
36. Li, X.; Hong, X.; Moilanen, A.; Huang, X.; Pfister, T.; Zhao, G.; Pietikainen, M. Reading Hidden Emotions: Spontaneous Micro-expression Spotting and Recognition. *arXiv* **2015**, arXiv:1511.00423.
37. Wright, J.; Ganesh, A.; Rao, S.; Peng, Y.; Ma, Y. Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices via Convex Optimization. In *Advances in Neural Information Processing Systems 22*; Bengio, Y., Schuurmans, D., Lafferty, J.D., Williams, C.K.I., Culotta, A., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2009; pp. 2080–2088.
38. Wang, S.J.; Yan, W.J.; Zhao, G.; Fu, X.; Zhou, C.G. Micro-Expression Recognition Using Robust Principal Component Analysis and Local Spatiotemporal Directional Features. In *Workshop at the European Conference on Computer Vision*; Springer: Cham, Switzerland, 2014; pp. 325–338.
39. Wang, S.J.; Yan, W.J.; Li, X.; Zhao, G.; Zhou, C.G.; Fu, X.; Yang, M.; Tao, J. Micro-Expression Recognition Using Color Spaces. *IEEE Trans. Image Process.* **2015**, *24*, 6034–6047. [[CrossRef](#)] [[PubMed](#)]
40. Huang, X.; Zhao, G.; Hong, X.; Zheng, W.; Pietikainen, M. Spontaneous facial micro-expression analysis using Spatiotemporal Completed Local Quantized Patterns. *Neurocomputing* **2016**, *175*, 564–578. [[CrossRef](#)]
41. Wang, S.J.; Yan, W.J.; Sun, T.; Zhao, G.; Fu, X. Sparse tensor canonical correlation analysis for micro-expression recognition. *Neurocomputing* **2016**, *214*, 218–232. [[CrossRef](#)]
42. Liong, S.T.; See, J.; Phan, R.C.W.; Wong, K. Less is More: Micro-expression Recognition from Video using Apex Frame. *arXiv* **2016**, arXiv:1606.01721.
43. Xu, F.; Zhang, J.; Wang, J.Z. Microexpression Identification and Categorization Using a Facial Dynamics Map. *IEEE Trans. Affect. Comput.* **2017**, *8*, 254–267. [[CrossRef](#)]
44. Wang, S.J.; Wu, S.; Qian, X.; Li, J.; Fu, X. A main directional maximal difference analysis for spotting facial movements from long-term videos. *Neurocomputing* **2017**, *230*, 382–389. [[CrossRef](#)]
45. Ekman, P.; Friesen, W.V. Measuring facial movement. *Environ. Psychol. Nonverbal Behav.* **1976**, *1*, 56–75. [[CrossRef](#)]
46. Platt, J.C. Fast training of support vector machines using sequential minimal optimization. In *Advances in Kernel Methods*; The MIT Press: Cambridge, UK, 1999; pp. 185–208.
47. Davison, A.K.; Lansley, C.; Ng, C.C.; Tan, K.; Yap, M.H. Objective Micro-Facial Movement Detection Using FACS-Based Regions and Baseline Evaluation. *arXiv* **2016**, arXiv:1612.05038.
48. Ng, C.C.; Yap, M.H.; Costen, N.; Li, B. Wrinkle detection using hessian line tracking. *IEEE Access* **2015**, *3*, 1079–1088. [[CrossRef](#)]
49. Wang, Y.; See, J.; Phan, R.C.W.; Oh, Y.H. Efficient Spatio-Temporal Local Binary Patterns for Spontaneous Facial Micro-Expression Recognition. *PLoS ONE* **2015**, *10*, e0124674. [[CrossRef](#)] [[PubMed](#)]
50. Bengio, Y. Learning Deep Architectures for AI. *Found. Trends Mach. Learn.* **2009**, *2*, 1–127. [[CrossRef](#)]
51. Deng, L.; Yu, D. Deep Learning: Methods and Applications. *Found. Trends Signal Process.* **2014**, *7*, 197–387. [[CrossRef](#)]

52. Alarifi, J.; Goyal, M.; Davison, A.; Dancey, D.; Khan, R.; Yap, M.H. Facial Skin Classification Using Convolutional Neural Networks. In Proceedings of the 14th International Conference on Image Analysis and Recognition, ICIAR 2017, Montreal, QC, Canada, 5–7 July 2017; Springer: Cham, Switzerland, 2017; Volume 10317, p. 479.
53. Soomro, K.; Zamir, A.R.; Shah, M. UCF101: A Dataset of 101 Human Actions Classes From Videos in The Wild. *arXiv* **2012**, arXiv:1212.0402.
54. Karpathy, A.; Toderici, G.; Shetty, S.; Leung, T.; Sukthankar, R.; Fei-Fei, L. Large-scale Video Classification with Convolutional Neural Networks. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).