Emotion Categorization from Video-frame Images using a Novel Sequential Voting Technique

Harisu Abdullahi Shehu[1][0000−0002−9689−3290], Will Browne[1][0000−0001−8979−2224], and Hedwig Eisenbarth[2][0000−0002−0521−2630]

[1] School of Engineering and Computer Science, Victoria University of Wellington, 6012 Wellington, New Zealand
{harisushehu,will.browne}@ecs.vuw.ac.nz
[2] School of Psychology, Victoria University of Wellington, 6012 Wellington, New Zealand
hedwig.eisenbarth@vuw.ac.nz

Abstract. Emotion categorization can be the process of identifying different emotions in humans based on their facial expressions. It requires time and sometimes it is hard for human classifiers to agree with each other about an emotion category of a facial expression. However, machine learning classifiers have done well in classifying different emotions and have widely been used in recent years to facilitate the task of emotion categorization. Much research on emotion video databases uses a few frames from when emotion is expressed at peak to classify emotion, which might not give a good classification accuracy when predicting frames where the emotion is less intense. In this paper, using the CK+ emotion dataset as an example, we use more frames to analyze emotion from mid and peak frame images and compared our results to a method using fewer peak frames. Furthermore, we propose an approach based on sequential voting and apply it to more frames of the CK+ database. Our approach resulted in up to 85.9% accuracy for the mid frames and overall accuracy of 96.5% for the CK+ database compared with the accuracy of 73.4% and 93.8% from existing techniques.

Keywords: CK+ · Emotion categorization · K-nearest neighbors · Random Forests · Sequential vote · Video-frame

1 Introduction

Significant effort has been made in developing emotion classification methods for human facial expressions. This is driven by the increasing number of intelligent systems where it is important to approximate an emotional state of mind, so as to improve their interaction with humans. Thus, emotion categorization becomes an increasingly important area of research in computer vision [1] [2] as classification from facial expression is so far the most readily available way to estimate states of emotion.

Emotion recognition is a common field of study. Here, we made the distinction between emotion recognition and emotion categorization as we contend that it is not possible to know the underlying emotion of a person because it can be superficially manipulated, e.g. bluffing in poker or business interactions.
Evidence has shown that human beings are not good at classifying emotion [3–5]. An image that will be classified to have a particular emotion by one person might be classified to have a different emotion by another person. This is due to the different perspectives we all have as a result of variation in neurobiological processes [6]. On the other hand, machine learning classifiers have done well in categorizing emotion when appropriate features are provided to the classifiers.
Researchers have used various methods to analyze emotion from posed and non-posed visual datasets[7] [8]. A posed dataset is generated by capturing the picture of participants in a controlled environment based on instructions given to them by an experimenter. Alternatively, datasets with non-posed expressions are created without instruction where labeling is post stimuli by "emotion experts" [1].

The CK+ database [9], which is mainly based on posed expressions, is chosen to be used as an example of an emotion video database. As the videos in the CK+ database have already been converted to image frames, this made it readily available to be used compared to other emotion video databases such as the DISFA [10] and FAMED [11] dataset, etc., in which the videos will need to be converted to image frames. Overall, the CK+ database has a total of seven classes of expressions, which comprise of the 6 basic (anger, disgust, fear, happiness, sadness, and surprise) emotions defined by Ekman [12] plus contempt expression.

The CK+ database does not label frames as "peak" i.e. where the emotion is most vividly expressed. However, the term peak is often applied to frame(s) in this dataset. A number of researchers use the last frame [13][14] whereas others use the last three frames as peak frames [15] [16]. Much research on the CK+ dataset uses a subset of the available frames where the emotion is considered at peak. This is anticipated to make the associated technique less effective in categorizing frames where the emotion is not expressed at peak.

The aim of this research is to use multiple frames of the CK+ database to analyze emotion labels from different image frames and compare our approach to that using just peak frames. Also to improve accuracy, a sequential voting technique that performs voting on each sequence of the video frames based on the prediction made will be applied.

The rest of the paper is organized as follows: Section 2 explains how people use video frame images for emotion categorization research using the CK+ database as an example. In Section 3, we explain the properties of the CK+ database and its emotion categories. In addition, we also explain our proposed method and introduce a sequential voting approach. Section 4 compares the use of more emotion video frames to fewer peak frames and shows the effectiveness of the sequential voting approach. In Section 5, we further discuss the obtained

[1] People that are trained in emotion categorization. These people labeled these databases based on the assumption that people smile when happy, frown their faces when sad, and scowl when anger irrespective of their age, race, and ethnicity.

result and highlight the shortcomings of the applied method. In Section 6, we conclude the paper and hint at certain limitations that could be addressed in future studies.

2 Related Work

Happy and Routray [17] proposed an approach based on Support Vector machine (SVM) multi-class classification. The face in each image is first detected followed by landmark detection of the region of interest such as the eyebrow corners, nose, eyes, and the lips corners. Also, active patch locations are defined with respect to the location of this landmark. All active patches are evaluated in the training phase and the ones with a maximum variation of features between the expressions are selected. A SVM multi-class classifier is used to classify the selected features after they are being projected into six different lower-dimensional (from 192 x 192 to 48 x 48 pixels) subspace on the CK+ database. The last images of every sequence of the six basic expressions where the expression is at its peak were selected resulting in a total of 329 images and the result was evaluated based on voting out two of the six different dimensions using 10-fold cross-validation, which lead to an accuracy of 94.09%. Although the CK+ has seven different facial expressions, Happy and Routray chose to analyze emotion from only six out of the seven expressions of the CK+ database as they consider the six basic expressions to be universal.

In contrast to [17], Elaiwat et al. [15] used all the seven expressions plus *neutral* provided by the CK+ database. They proposed an approach based on Restricted Boltzmann Machines (RBM) on the CK+ dataset in which voting is performed in the validation phase. Three image pairs were constructed from each of the 327 labeled sequences of the CK+ database in which the first image corresponded to neutral and the remaining two images corresponded to the strongest expression of that particular expression. Each image pair from the constructed pairs voted for one of the seven expressions and the expression class with the highest number of votes was considered to be the expression of the sequence. 10-fold cross-validation was used in the evaluation process, which leads to an accuracy of 95.66%. Surprisingly, the time it takes the method of Elaiwat et al. to train on a single epoch was significantly lower compared to the current-state-of-the-art approach as the training phase had been done off-line.

Similarly, Kim et al. [16] also used three frames from each sequence to analyze emotion labels from the CK+ database. They proposed an approach based on a hierarchical deep neural network. The first network performs feature extraction using a convolution neural network (CNN) whereas the second method extracts changes to the features and learns to identify all the six basic emotions. Adaptive weighing function is used to combine the result of the two features for the final result. Like [15], Kim et al. also used 10-fold cross-validation to evaluate their result and have achieved an accuracy of 96.46%. Unlike [17], the proposed method used dynamic features as opposed to static features and at the same time utilized a dual network instead of a single network.

Thus, all of the work by [15] [16] [17] used fewer frames than available of the CK+ database where the emotion is expressed at peak. However, a model trained with fewer frames where emotion is only expressed at peak might not perform well in recognizing frames where the emotion is not expressed at peak. Besides, peak expressions are rare in everyday life [18]. Therefore, we aim to address the issues of using fewer peak frames by using more frames of the CK+ database and compare the robustness of the two different approaches.

3 Methods

3.1 Dataset

In this research, we used the CK+ dataset [9] which is an extended version of the Cohn-Kanade database that was released in 2000 [19]. The CK+ was developed due to certain limitations on the CK dataset, including but not limited to non-validated emotion labels, lack of common performance metrics to evaluate new algorithms, and also the non-existence of a standard protocol for a common database.

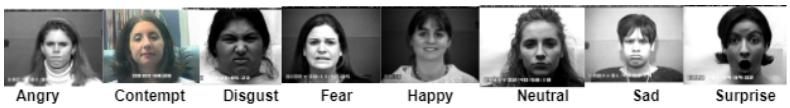

Fig. 1. Sample expressions from the CK+ database. Note that majority of the images from the videos are black and white

In the CK+ database, a sample of 201 adults between the ages of 18-50 years that comprises of 69% female, 13% Afro-American, 81% Euro-American, and 6% from other groups were recorded using AG-7500 cameras. A series of 23 facial displays were instructed to the participants by the experimenter. Certain participants smile to the experimenter between the task, which are also included in the dataset and as a result, the CK+ does not only contain posed but at the same time, few non-posed expressions. Fig. 1 shows examples of expressions from the CK+ database. As not all images in the CK+ database are labeled, only the labeled images are used. The image sequence in the CK+ varies from 10 to 60 frames starting from neutral to peak expressions and a total of 593 labeled sequences from 123 subjects.

3.2 Hardware specification

A Dell computer with Intel(R) Core(TM) i7-8700 CPU @ 3.2GHz processor, utilizing Windows 10 Education and 15.8 GB usable RAM is used in this research.

3.3 Proposed Methodology

There are several important factors such as resolution, illumination effects, and intensity of expressions to consider when classifying human facial expressions [20] [21]. They are considered to be important because they are the primary information stored within pixels.
Visual inspection of randomly selected frame sequences lead to the decision to use the second-half of the frames for training. Thus, in this research, the second-half of frames from each sequence of the CK+ database are assigned the emotion label of the sequence. In cases where the number of frames is odd, the value of the least succeeding integer is taken. For instance, we assign disgust to frames starting from 6 to 11 in a given sequence where the number of frames is 11 and the decoded emotion is disgust as illustrated in Figure 2.

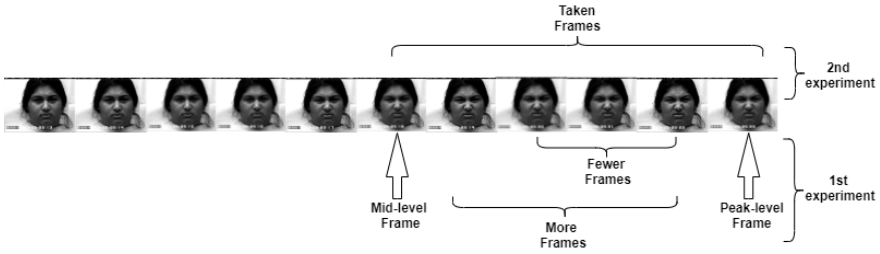

Fig. 2. Sample usage of the CK+ database. Note that the last-half of the frames of the CK+ are referred to as the taken frames, the first taken frame is referred to as the mid-level frame, the last taken frame is referred to as the peak-level frame, the last three frames preceding the peak-level frame are considered to be the fewer frames and finally, all frames between the mid-level frames and peak-level frames are referred as more frames of the CK+ database.

Figure 2 shows example changes of image frames of the CK+ database from neutral to peak expressions and also shows how the CK+ database is used in this research. Henceforth, we will refer to the first taken frame of each sequence where the emotion is not expressed at peak as the mid-level frame and the last taken frame of each sequence where the emotion is expressed at peak as the peak-level frame. As we aim to use more frames, all frames between the mid-level and peak-level frames are used for training a particular model.

To compare the advantage of using more frames over fewer frames, the last three frames before the peak-level frames of each sequence where the emotion is at peak are also used for training a separate model. After that, both models trained using more frames and fewer peak frames are tested with both mid-level and peak-level frames in the first experiment.

Fig. 3 represents the flow chart of the novel method. Blocks represented in the flow chart are explained in this section.

– Pre-processing: Emotion labels were converted to integers and image pixel values to an array. Normalization has also been performed on the pixels of the raw images as a pre-processing technique to normalize all pixel values between the range of [0, 1] to enable fast computation.
– Feature-extraction: Many machine learning algorithms can accomplish the task of image classification [22] [23] [24], however, all algorithms require proper features for conducting the classification. In this research, image color information is split into three different (RGB) channels as shown in Fig. 3.

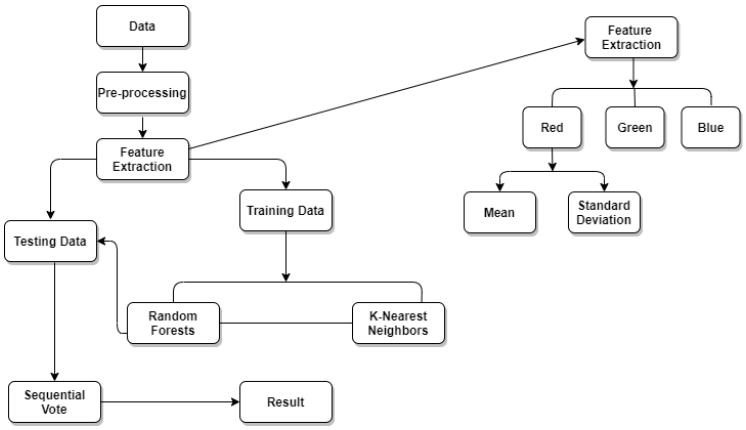

Fig. 3. Flow of the novel method.

Furthermore, since most of the video frames of the CK+ database are in grayscale, we know that all the RGB channels should have the same value. Therefore, to avoid having redundant features, two feature vectors of only the red (R) channel are extracted from each image. The first feature is the mean whereas the second feature is the standard deviation of the R channel. These features are later used for the classification of the images.
– Classification: After feature extraction has been performed on the data, both features extracted from mid-level and peak-level frames are classified by models train on more and fewer peak frames. Thereafter, in order to reflect generalized performance on the data, stratified 10-fold cross-validation is used to test the performance of the model before applying the voting technique.

3.4 Machine Learning Classifiers

Random Forest (RF) and K-Nearest Neighbour (KNN) are used in this research. RF is chosen to be used in this research because it is an ensemble technique that is averaged over many trees. Therefore, it has a higher chance of achieving a higher classification accuracy compared to other machine learning classification techniques such as decision tree (DT) that explore fewer decision boundaries

[25]. KNN is chosen to be used due to its simplicity as it requires no explicit training phase. Another reason why the algorithms are chosen to be used is based on their promising performance obtained in the research of [26] and [27] respectively.

Random forest is an extension of the decision tree (DT) algorithm [28], which uses control conditional statements to predict an outcome. It constructs multiple DTs during training and merges them into a single forest [29] [30]. The goal is to rely on a collection of decisions from the multiple constructed DTs to improve accuracy [31]. The algorithm does not allow overfitting trees in the model, so maintains the prediction accuracy over a large proportion of data.

The K-nearest neighbor is often referred to as a lazy learner because it does not learn from the training data. In KNN, objects are classified based on the plurality vote of their neighbors [32].

3.5 Sequential Vote

The sequential voting (SV) process is a conditional statement, which is performed based on what the majority of images in a given sequence are predicted to be. Since the CK+ database is a database of mainly posed video converted to image frames, it is assumed that all images in a single video sequence belong to a single class. Label flickering is a common problem that occurs in video classification. It is an unusual change from the actual frame label to a different frame label that occurs when predicting the same sequence of video frames. The sequential voting is performed to reduce the possibility of label flickering occurring.

Algorithm 1 shows the procedure of how SV is performed on a single sequence predicted by a classifier. If the mode (i.e. the most common label) of a particular sequence is more than one, the sequence label remains unchanged. Otherwise, all images in that sequence are assigned the same label as the mode.

Algorithm 1 Sequential Voting Algorithm.

Get image sequence
Get predictions from algorithm
Begin
For each sequence i
Find *mode* of predicted sequence_{label}
if $\text{len}(\text{sequence}_{mode}) > 1$ **then**
    continue
end if
if $\text{len}(\text{sequence}_{mode}) = 1$ **then**
    $\text{sequence}_{label} = mode$
end if
    return sequence_{label}

Most research on the CK+ database is performed either on the six basic emotions or six basic emotions plus contempt expression provided by the CK+

database. For the sake of comparison, results are presented using both six basic expressions and also 6 basic plus contempt expression provided by the CK+ database.

4 Results

Table 1 shows the percentage accuracy obtained after testing a model trained with fewer frames with mid-level and peak-level frames of the CK+ database. The overall accuracy of 72.5% has been achieved when tested with mid-level frames whereas an accuracy of 93.3% has been achieved when tested with peak-level frames.

While the model trained with fewer frames is not able to achieve an accuracy of more than 83% on any of the classes when tested with mid-level frames, an accuracy of not less than 89% has been achieved on all classes when the same model is tested with peak-level images.

A k value of one is used for the KNN algorithm and the algorithm is deterministic on the CK+ database. Unlike KNN, as the RF algorithm is not deterministic, results achieved by the RF algorithm are presented with upper and lower bound with a 95% confidence interval.

* refers to mean accuracy across all categories in Table 1 and 2. In the expression section of Table 5, *An* represents anger, *Di* represents disgust, *Fe* represents fear, *Ha* represents happy, *Sa* represents sadness and *Su* represents surprise expression.

Table 1. Percentage of correctly classified classes by model trained on fewer frames and test with mid and peak level images using KNN

Testpoint	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Mid-level	79	75	68	83	71	59
Peak-level	100	90	92	96	89	93

*$Mid-level$ =72.5%, Peak-level = 93.3%

Table 2 shows the percentage accuracy obtained after testing a model trained with more frames with mid-level and peak-level frames of the CK+ database respectively. An accuracy of 85.8% has been achieved when tested with mid-level frames whereas the achieved accuracy is up to 92% when the model is tested with peak-level images.

Although the accuracy achieved by the model trained with more frames is lower when tested with mid-level frames compared with the achieved accuracy on peak-level frames, the achieved accuracy is up to 13.3% higher than the accuracy achieved when the same frames are predicted by the model trained with fewer peak frames of the CK+ database.

In addition, the accuracy achieved in all the classes in Table 2 when predictions are made on the mid-level frames by a model trained with more frames is

Table 2. Percentage of correctly classified classes by model trained on more frames and test with mid and peak level images using KNN

Testpoint	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Mid-level	93	86	88	93	82	73
Peak-level	98	86	84	94	100	90

*$Mid-level$ =85.8%, Peak-level = 92%

higher than the accuracy achieved in all the classes in Table 1 when the same mid-level frames are predicted by a model trained with fewer peak frames.

Table 3. Accuracy obtained from predictions on mid-level and peak-level images made by model trained on more and fewer frames

Algorithm	**Classes**	**No. of training images**	**Testpoint**	**Accuracy**
RF	6 basic	927	mid-level	68.9±0.3
			peak-level	90.2±0.3
KNN			mid-level	72.5
			peak-level	93.3
RF	6 basic + contempt	981	mid-level	70.5±0.3
			peak-level	90.6±0.2
KNN			mid-level	73.4
			peak-level	93.8
RF	6 basic	2132	mid-level	79.9±0.2
			peak-level	90.0±0.3
KNN			mid-level	85.5
			peak-level	92.0
RF	6 basic + contempt	2205	mid-level	81.1±0.2
			peak-level	89.5±0.2
KNN			mid-level	85.9
			peak-level	92.4

Table 3 shows the accuracy achieved by KNN and RF when the prediction is made on mid-level and peak-level images by both model trained with more and fewer frames. While the models trained on fewer frames are trained with less than a thousand images, more than two thousand images are used to train the model with more frames. Based on these results, the accuracy achieved by KNN is higher than the accuracy achieved by the RF algorithm in all of the cases.

From the results obtained, it can be seen that the use of fewer frames achieved only a slightly better accuracy when predictions are made on peak-level frames. However, a linear regression analysis predicting accuracy by *Testpoint* and *Framesize* resulted in a significant difference.

Table 4 presents the regression results of the statistical test performed. The test is performed after getting the result of 30 runs of each case when the mid-

level and peak-level frames are classified by two different classifiers trained on fewer and more frames of the CK+ database. The test is performed across 120 samples using the *Least Squares* method. The *Accuracy* is used as the dependent variable across *Framesize* and *Testpoint* as independent variables.
Framesize species whether the model is trained with more or fewer frames and *Testpoint* indicates whether the classier is used to classify mid-level or peak-level frames.

Table 4. Regression Results

	Coefficient	**P**	**[0.025**	**0.975]**
Intercept	0.6889	<.001	0.687	0.691
C(Testpoint)	0.2127	<.001	0.209	0.216
C(Framesize)	0.1079	<.001	0.105	0.111
C(Testpoint):C(Framesize)	-0.1207	<.001	-0.125	-0.116
Adjusted R^2	0.994			
F-statistics	6769			
AIC	-862.0	**BIC**	-850.8	

As can be seen from Table 4, a p-value of p <.001 of the t-statistic of the Ordinary Least Squares (OLS) method shows that the result obtained by the model trained with more frames was significantly better than the result obtained by the model trained with fewer frames specifically in the case of predicting the mid frames. Also, the overall *effect size* of 0.553 has been found for the factor Framesize.
These results suggest that the use of more frames provides a better performance compared wıth the use of fewer peak frames. Consequently, in order to apply sequential voting, we assigned all second-half (taken) frames of each sequence the emotion label of the sequence (see Fig. 2) in the second experiment.

Table 5 shows the accuracy achieved by RF and KNN from six basic emotions both before and after the sequential voting is performed. It can be seen clearly from the table that the result obtained after the sequential voting process increases the overall accuracy by 6% and 5.5% in RF and KNN algorithms respectively.
The accuracy in each class has also increased after the sequential voting. The sad class has seen the most increase with up to 8%, from 92% to 100%. This happens because most of the image frames in the sad class sequence are predicted correctly by the classifier and as a result, the sad class sequence always appears to have a single mode which the voting algorithm uses to correctly change the label of the wrongly predicted frames.

Table 6 shows results achieved on CK+ database both before and after SV using KNN and RF algorithms. Surprisingly, before the SV is performed, the accuracy achieved by KNN is higher than the accuracy achieved by RF on both six basic and six basic plus contempt expression. However, after the SV, the result

Table 5. Prediction accuracy on six basic emotions of the CK+ database before and after sequential voting (+SV)

Expressions	Algorithms			
	RF	RF + SV	KNN	KNN + SV
An	94	99	94	98
Di	88	95	90	93
Fe	90	96	92	95
Ha	95	98	95	98
Sa	92	100	93	99
Su	84	91	85	93
Standard dev.	0.18	0.21	0.0	0.0
Average acc.	90.5±0.1	96.5±0.1	91.5	96.0

achieved by RF surpasses the result achieved by KNN on six basic emotions and equivalent to the result achieved by KNN on six basic plus contempt emotional expression.

Table 6. Sequential voting result from the CK+ database

Method	**Classes**	**Accuracy**
RF	6 basic	90.5±0.1
RF + SV	6 basic	96.5±0.1
KNN	6 basic	91.5
KNN + SV	6 basic	96.0
RF	6 basic + contempt	90.6±0.1
RF + SV	6 basic + contempt	96.2±0.1
KNN	6 basic + contempt	91.5
KNN + SV	6 basic + contempt	96.2

5 Discussion

This study set out with the aim of assessing the importance of using more frames of an emotion video database to categorize emotion from different image frames.

In reviewing the literature, many studies [15] [16] [17] are found to be using fewer frames of the CK+ database where the expression is at peak to predict emotion labels. Several other studies [13][14] recently conducted on the CK+ database using the state-of-the-art deep learning algorithms were also found to be using fewer (typically final frames) than the available frames of the CK+ database. While a high accuracy result is obtained on peak-level frames when fewer peak frames are used to train the model, the results were not very encouraging when the same model is used to predict emotion labels from mid-level

frames. To the best of our knowledge, this is the first study to use more varied frames of sequential video images to predict emotion.

The model trained with more frames achieved an accuracy of up to 1.3% and 1.4% lower than the accuracy achieved by the model trained with fewer peak frames when the prediction is made for peak-level frames on the 6 basic and 6 basic plus contempt expression. However, the overall performance of the model is significantly better than that of the model trained with fewer frames.

RF algorithm has seen the strongest increase in accuracy when SV is applied in comparison to the result the algorithm achieved as SV increase the result with up to 6% and 5.6% compared to the increase of 4.5% and 4.7% on 6 basic and 6 basic plus contempt expression when the same SV technique is applied to KNN algorithm.

Overall, these results indicate that the application of SV to reduce label flickering increases the categorization accuracy achieved by both RF and KNN algorithms.

Several reports [33] [34] have shown that the state-of-the-art deep learning algorithms take a very long time, ranging from hours to weeks to train on facial expression datasets even using Graphical Processing Units (GPU). We also know from our previous work [26] that deep learning methods such as the residual neural network (ResNet) takes over an hour to train on the CK+ database. Thus, compared to these state-of-the-art approaches, the evidence presented thus far supports the idea that our approach achieved a considerably lower execution time. It takes between 5.0 ± 0.04 to 5.4 ± 0.01 seconds to compute the result of the CK+ database when tested on both more frames and the sequential voting technique respectively.

Based on the results obtained, we can say that here, we have set a sub-standard for other researchers on how to use video-frames with facial expressions to perform emotion classification research based on emotion labels.

6 Conclusion and Future Work

This paper proposed an approach to use more, over fewer, peak frames where emotion is expressed at peak to analyze emotion classes in the CK+ database. We have compared our approach to an approach using fewer peak frames and have achieved a better accuracy result when the prediction is made on mid-level images. We have also found that the use of more frames to train the model gives a significantly better performance compared to when fewer peak frames are used. Furthermore, we have shown that performing sequential voting on the results obtained by RF and KNN classifiers increases the accuracy further.

This study is carried out on posed emotion video frames images of the CK+ database and therefore despite these promising results, questions remain on whether the same technique could be used on non-posed emotion video-frames as well as on other datasets. Future work should, therefore, apply this new approach to non-posed and other emotion video datasets.

References

1. Tian, Y.-L., Kanade, T., & Cohn, J. F., (2005) "Facial expression analysis," in Springer: Handbook of face recognition, pp. 247-275.
2. Martinez, B. & Valstar, M. F. (2016) "Advances, challenges, and opportunities in automatic facial expression recognition," in Springer: Advances in Face Detection and Facial Image Analysis, pp. 63-100.
3. Matsumoto, D., & Hwang, H. S. (2011). Evidence for training the ability to read microexpressions of emotion. Motivation and Emotion, 35(2), 181–191.
4. Krumhuber, E. G., Küster, D., Namba, S., Shah, D., & Calvo, M. G. (2019). Emotion recognition from posed and spontaneous dynamic expressions: Human observers versus machine analysis. Emotion.
5. Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. Psychological Science in the Public Interest, 20(1), 1–68.
6. Chakravarti, A. (2015). Perspectives on human variation through the lens of diversity and race. Cold Spring Harbor Perspectives in Biology, 7(9).
7. Islam, B., Mahmud, F., & Hossain, A. (2019). Facial Region Segmentation Based Emotion Recognition Using Extreme Learning Machine. 2018 International Conference on Advancement in Electrical and Electronic Engineering, ICAEEE, 1–4.
8. Mahmud, F., Islam, B., Hossain, A., & Goala, P. B. (2019). Facial Region Segmentation Based Emotion Recognition Using K-Nearest Neighbors. 2018 International Conference on Innovation in Engineering and Technology, ICIET 2018, 1–5.
9. Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010, 94–101.
10. Mavadati, S. M., Mahoor, M. H., Bartlett, K., Trinh, P., & Cohn, J. F. (2013). Disfa: A spontaneous facial action intensity database. IEEE Transactions on Affective Computing, 4(2), 151-160.
11. Longmore, C. A., & Tree, J. J. (2013). Motion as a cue to face recognition: Evidence from congenital prosopagnosia. Neuropsychologia, 51, 864-875
12. Ekman, P. & Friesen, W. V., (1971). Constants across cultures in the face and emotion. Journal of personality and social psychology, vol. 17, no. 2, p. 124.
13. A. Mollahosseini, D. Chan & M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, 2016, pp. 1-10
14. Minaee, S., & Abdolrashidi, A. (2019). Deep-emotion: Facial expression recognition using attentional convolutional network. arXiv preprint arXiv:1902.01019.
15. Elaiwat, S., Bennamoun, M., & Boussaid, F. (2016). A spatio-temporal RBM-based model for facial expression recognition. Pattern Recognition, 49, 152–161.
16. Kim, J. H., Kim, B. G., Roy, P. P., & Jeong, D. M. (2019). Efficient facial expression recognition algorithm based on hierarchical deep neural network structure. IEEE Access, 7, 41273–41285.
17. Happy, S. L., & Routray, A. (2015). Automatic facial expression recognition using features of salient facial patches. IEEE Transactions on Affective Computing, 6(1), 1–12.
18. Ruiqi, X., Xianchun, L., Lin, L., & Yanmei, W. (2016). Can We Distinguish Emotions from Faces? Investigation of Implicit and Explicit Processes of Peak Facial Expressions. Frontiers in Psychology 7:(1664-1078) pg 1330.

19. Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. Proceedings - 4th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2000, (March), 46–53.
20. Ting, G., Moydin, K., & Hamdulla, A. (2018). An Overview of Feature Extraction Methods for Handwritten Image Retrieval. Proceedings - 2018 3rd International Conference on Smart City and Systems Engineering, ICSCSE 2018, 840–843.
21. Pisal, A., Sor, R., & Kinage, K. S. (2018). Facial Feature Extraction Using Hierarchical MAX(HMAX) Method. 2017 International Conference on Computing, Communication, Control and Automation, ICCUBEA 2017, (figure 2), 1–5.
22. Loussaief, S., & Abdelkrim, A. (2017). Machine learning framework for image classification. 2016 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications, SETIT 2016, 58–61.
23. Li, Y., Wang, S., Zhao, Y., & Ji, Q. (2013). Simultaneous facial feature tracking and facial expression recognition. IEEE Transactions on Image Processing, 22(7), 2559–2573.
24. Cruz, A. C., Bhanu, B., & Thakoor, N. S. (2014). One shot emotion scores for facial emotion recognition. 2014 IEEE International Conference on Image Processing, ICIP 2014, (C), 1376–1380.
25. Safavian, S. R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. IEEE transactions on systems, man, and cybernetics, 21(3), 660-674.
26. Shehu H. A., Browne W., & Eisenbarth H. (2020). An Adversarial Attacks Resistance-based Approach to Emotion Recognition from Images using Facial Landmarks. 2020 IEEE International Conference on Robot and Human Interactive Communication.
27. Sohail, A. S. M., & Bhattacharya, P. (2007, March). Classification of facial expressions using k-nearest neighbor classifier. In International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications (pp. 555-566). Springer, Berlin, Heidelberg.
28. Kaminski, B., Jakubczyk, M., & Szufel, P. (2017). "A framework for sensitivity analysis of decision trees". Central European Journal of Operations Research. 26 (1): 135–159.
29. Shehu, H. A., Tokat, S., Sharif, M. H., & Uyaver, S. (2019, December). Sentiment analysis of Turkish Twitter data. In AIP Conference Proceedings (Vol. 2183, No. 1, p. 080004). AIP Publishing LLC.
30. Shehu, H. A., & Tokat, S. (2019, April). A hybrid approach for the sentiment analysis of Turkish Twitter data. In The International Conference on Artificial Intelligence and Applied Mathematics in Engineering (pp. 182-190). Springer, Cham.
31. Breiman L. (2001). "Random Forests". Machine Learning. 45 (1): 5–32.
32. Altman, N. S. (1992). "An introduction to kernel and nearest-neighbor nonparametric regression". The American Statistician. 46 (3): 175–185.
33. Fan, Y., Lam, J. C., & Li, V. O. (2018, October). Multi-region ensemble convolutional neural network for facial expression recognition. In International Conference on Artificial Neural Networks (pp. 84-94). Springer, Cham.
34. Chengeta, K., & Viriri, S. (2019, March). A review of local, holistic and deep learning approaches in facial expressions Recognition. In 2019 Conference on Information Communications Technology and Society (ICTAS) (pp. 1-7). IEEE.