Automated Facial Expression Recognition using SVM and CNN

Tadese Henok Seifu

Tianjin University of Technology and Education,
School of Information Technology and Education, 1310 Dagu South Road,
Hexi District, Tianjin, PRC – 300222, China

Abstract:- **Automatic emotion recognition via facial expressions is a fascinating area of study, which is used in a variety of fields, including safety, health, and human-machine interactions. In the fields of computer vision and artificial intelligence, facial emotion recognition plays a critical role. Furthermore, in order to make human-computer interaction effective in advancing artificial intelligence and humanistic robotic applications, real-time facial recognition programs must be able to execute at faster speeds and accurate rates. This work utilizes the Jaffe database. The Viola-Jones algorithm is used to recognize faces. Then image preprocessing is performed. PCA approach is used to extract features from the facial images. Then SVM and CNN classifiers are used to classify features such as fear, anger, disgust, happy, sad and surprise. Both the classifiers performed well on the dataset.**

Keywords: Automated facial expression recognition · Viola-Jones algorithm, CNN, SVM

1. INTRODUCTION

Hand, voice, body gestures, and facial expressions are the most common ways to express emotions. During conversations, facial expressions are utilized to convey feelings. Only 55% of emotions are communicated by facial expressions [1]. Six basic universal emotion expressions have been found by [2]. [3] Conducted a comprehensive study on facial emotion analysis a few decades ago and identified six primary expressions: anger, pleasure, sadness, disgust, surprise, and fear. Necessary information cues are displayed on the human face to convey emotional experience or conduct. In just a few seconds, humans can effectively discern a person's emotions by examining their face. Human–computer interaction [4], clinical services, and pupil knowledge assessment [5], multimedia, expression enabled equipment [6], surveillance [7], autism illness patients [8], and driving safety [9, 10] are all applications of facial emotion recognition.

JAFFE, which is a challenging dataset due to its inter-class and intra-class similarities, has been employed in this research to recognize facial emotions. Inter-class similarity occurs when certain images from various expression classes have similar appearances, making discrimination difficult. Intra-class variation occurs when certain images in the same expression class exhibit various variables such as illumination, age, and skin tone, making it harder for the model to distinguish the expression. For face expression recognition, intra-class variances are impossible to overcome. Due to substantial intra-class differences and high inter-class resemblance caused by minor face appearance changes, illumination changes, skin-color changes, and identity-related variables such as age, gender, and race, FER's performance declines in virtual environments.

Many studies were carried out in the literature on datasets with relatively minor intra-class differences. When we distinguish virtual facial emotions from virtual situations, however, the condition is difficult to meet. To address the aforementioned issues, researchers have offered a variety of techniques. Several traditional systems to FER, on the other hand, did not explicitly incorporate intra-class variation, although they did use datasets with intra-class variability. To accomplish virtual characters expression recognition, the majority of previous approaches [11, 12, 13] rely on designed characteristics with limited generalization capacity.

Recent advances in computer vision, particularly deep learning models, have increased face emotion classification performance. Convolutional Neural Network (CNN)-based models are extremely reliable and perform well in facial expression classification. CNN fine-tunes convolutional filter parameters at each layer to provide high-level features that adapt and describe the desired attributes for classifying unidentified images.

[14] Solve the intra-class variation problem by using training images to create an intra-class variant image for each expression, and the differences between these images are the attributes for dimension reduction. The lighting variation is addressed by this strategy. To reduce the intra-class variation influence, [15, 16] employed intra-class variation reduction features. The impacts of skin color and age fluctuations were not taken into account in this procedure.

Inter-class similarity and intra-class variation hampered the performance of existing FER systems. To overcome these concerns, this work develops a model for extracting discriminative features from virtual characters' faces in order to distinguish their seven facial emotions. The model extracts discriminative features from the virtual characters using the PCA technique. The discriminative capacity of features can help to lessen the influence of intra-class and inter-class similarities, making the model more robust in the face of changes. SVM and CNN classifiers have been used to recognize these features.

2. RELATED WORK

Facial expression recognition and classification have long been regarded as a difficult problem in emotional analysis. Numerous deep learning and machine learning (ML) models for emotion recognition tasks have been presented and developed by many authors in recent years. The intra-class variation was not explicitly considered in most previous publications, but they conducted experiments utilizing datasets with intra-class variability.

Using Gabor wavelets and DCT (Discrete Cosine Transform), [17] suggested a fusion-based technique for

identifying emotions. Using Gabor filters and DCT, a new form of the feature was retrieved in this study. To extract features and minimize dimensionality, the kernel principal component analysis was used. To categorize the expression images into six basic emotions, the RBFNN (radial basis function neural network) was used. The CK dataset was used in the experiments, and an accuracy of 99 percent was achieved with only a few training and testing samples. [18] Used the Supervised Committee of CNNs to create a framework for recognizing emotions. For feature extraction, 72 CNNs with identical baseline architecture were used. On the FER2013, MMI, and LFW datasets, the proposed work was evaluated. [19, 20] used Attention Mechanism (ACNN) to create a CNN model for recognizing emotions. pACNN focused on local facial patches, while gACNN took into account both patch-level and image-level features. On the Affect Net and RAF-DB datasets, experimentation yielded 85 and 58.75 percent accuracy, respectively.

[15, 16] created a Deep Comprehensive Multi patches Aggregation CNNs-based model. Two CNN branches were used in this study. One branch of CNNs was used to extract local features from patches, while the other branch was used to extract holistic characteristics from the full-face sample. These features were then combined to form a feature vector, which was then passed to the classifier for expression categorization. Experiments on the CK+ and JAFFE datasets yielded an accuracy of 93.46 and 94.75 percent, respectively. [11] Used DCNNs to build a new method for identifying emotions. The first face was discovered from dataset photos, and those frontal face images were passed to CNN for feature extraction. The classification was done using SVM with a grid search. The proposed models were tested on CK+ and JAFEE and scored 97 and 98.12% accuracy, respectively. Different FER approaches were proposed by [10, 12]. Gabor wavelets and HWT were used to extract local and global features in this study. The feature dimension was reduced using non-linear PCA (NLPCA). To merge those

two types of characteristics, weighted and concatenated fusion techniques were used. SVM was used to classify the data. The CK+ was used in the experiment, and it was shown to be 98 percent accurate.

An RGB–D Microsoft Kinect camera was modified to record pupils' facial expressions in the classroom in order to recognize emotions [21]. To train and categorize the expressions, researchers employed the Adaptive-Network-Based Fuzzy Inference System machine learning technique. The system was trained using a combination of the EURECOM and Cohn-Kanade datasets. The quality of the supplied photos determines the accuracy of biometric recognition systems. In [22], the impact of image quality on accuracy was examined. In this investigation, the system provided good accuracy till the raw picture compression ratio of 30–40% and greater ratios had a detrimental impact on the system's accuracy. To extract the richer characteristics from macro pixels, [19, 20] included deep overlap and weighted filter principles into the macro pixel technique. The experiment results reveal that the proposed strategy outperformed the original macro pixel approaches in terms of accuracy.

Connie et al. combine CNN features with SIFT features to improve the FER accuracy [23]. The correctness of this work was tested using the FER2013 and CK+ datasets, and it was found to be 73.4 percent and 99.1 percent, respectively. [24] Created a CNN feature-based FER in which facial characteristics were extracted using CNN. Mixing diverse dataset photos increased model generalization. For FER on static images, [25] employ transfer learning with hyperparameter optimization. To improve the model's accuracy, they used hyperparameter optimization. The JAFFE and ERUFER datasets were used to test this work. With improved execution speed, [13] established a collaborative optimization framework for FER employing local binary features and shallow networks. [26] Created a hybrid deep learning model for facial emotion recognition. One CNN classified the primary emotion as sad or happy, while a secondary CNN recognized the image's secondary emotion. The FER2013 and JAFFE datasets were used to test this study.

All of the aforementioned studies produced positive results on human-based datasets, however, these models are sensitive to the lighting and specific poses present in the dataset. Because of the aforementioned issues, the performance of existing FER systems was limited. A single classifier was employed in the majority of current algorithms for recognizing facial expressions. As a result, a new model that recognizes emotions more accurately has a lot of potentials.

3. METHODOLOGY

The proposed model utilizes the JAFFE database. Data augmentation is used to enhance the number of images. Then the faces are recognized and cropped. After this, preprocessing is performed on the faces. These faces are sent into PCA for feature extraction. Finally, SVM and CNN classifiers are used to recognize features.
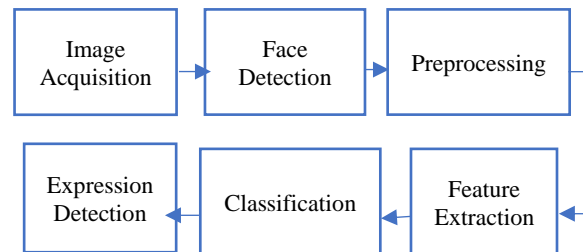

Figure 1: Proposed model of the system

4. PROPOSED NETWORK ARCHITECTURES

4.1 Data Augmentation

Deep learning models demand more samples for training. JAFFE dataset has limited samples and there is a risk of underfitting, so image data augmentation is used to expand the dataset. Data augmentation techniques such as adding gaussian noise, rotation, and shifting are used.

4.2 Data Preprocessing

Image preprocessing techniques such as normalization, resizing and rotation correction are applied to the faces.

4.3 Training

The system utilizes 90% data for training and 10% data for testing. Facial images are sent into the system during training and the network weights are learned.

4.4 Testing

The system gets facial images during testing and finally displays the classifier's accuracy.

5. EXPERIMENTS

5.1 Dataset

JAFFE database is used in the proposed model. Data augmentation is used to enhance the number of images. Figure 1 shows some images of the JAFFE dataset.

Figure 2: JAFFE dataset images

5.2 Experiment Setup

The proposed model is tested on a system with an Intel Core i5 processor, 8 GB of RAM, and a 500 GB hard drive. Coding is done in MATLAB 2019a.

5.3 Results and Evaluation

5.3.1 The affection of data augmentation

By creating new and varied samples for training datasets, data augmentation has enhanced the efficiency of machine learning models.

5.3.2 The affection of test method

The system used 90% data for training and 10% data for testing and showed good results.

5.3.3 The affection of loss function

Machine learning algorithms are greatly improved and recognized via loss functions.

5.3.4 Comparison between different models

Both the models performed well on the given dataset. The performance of SVM classifier is shown in figure 3.


```
Editor - G:\Freelancing Projects\Research Papers\4. Automated Facial expression recognition us
  SVMCode.m   ×   +
1
2        %% Training Code
3 -      featureCount = 1;
4 -      rootFolder = fullfile('Training Images');
5 -      TrainingimgSets = [imageSet(fullfile(rootFolder,'0')),...
6            imageSet(fullfile(rootFolder,'1')),...
7            imageSet(fullfile(rootFolder,'2')),...
8            imageSet(fullfile(rootFolder,'3')),...
9            imageSet(fullfile(rootFolder,'4')),...
10           imageSet(fullfile(rootFolder,'5'))];
Command Window
New to MATLAB? See resources for Getting Started.
  >> SVMCode

  SVMAccuracy =

      75
```
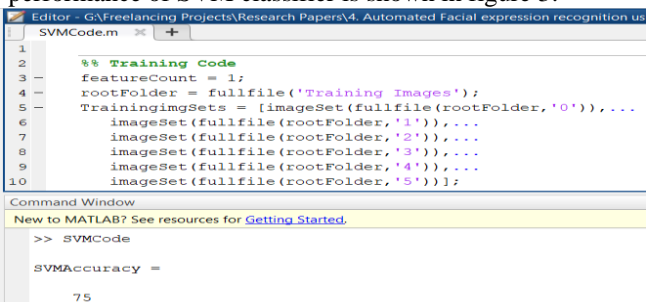
Figure 3: Performance of SVM Classifier

The performance of CNN classifier is shown in figure . The results show that CNN shows good performance when large amount of data is used.


```
Editor - G:\Freelancing Projects\Research Papers\4. Automated Facial expression recognition using svm and c
  MainCode.m   ×   +
1
2        %% Training Code
3 -      imds = imageDatastore('Images','IncludeSubfolders',true,...
4            'LabelSource','foldernames');
5 -      [imdsTrain,imdsTest] = splitEachLabel(imds,0.90,0.10,'randomize');
6 -      layers = [
7            imageInputLayer([256 256])
8            convolution2dLayer(3,8,'Padding','same')
9            batchNormalizationLayer
10           reluLayer
Command Window
New to MATLAB? See resources for Getting Started.
  >> MainCode

  TestingAccuracy =

      98.3607
```

Figure 4: Performance of CNN Classifier

6. CONCLUSION

The purpose of this research has been to address the field of facial expression recognition. This subject of study has been thoroughly studied in terms of applicability and automation, beginning with the psychological reason for face behavior analysis. Psychologists' manual face analysis was swiftly supplanted by appropriate computer software. To suit the requirements of the facial expression recognition system, a wide range of image processing algorithms were developed. However, numerous obstacles and problems remain with such systems, particularly in terms of improving their performance and application.

This work includes the design and implementation of a Facial Expression Recognition System, in addition to the theoretical background. The proposed system was developed to process facial behavior images and recognize behaviors in terms of six basic emotions. Full automation, as well as user and environment independence, are key features of the system. Occlusions are not a problem for the system. Furthermore, the results of the recognition are extremely encouraging. For the given dataset, the SVM classifier has performed better than CNN.

In the future, I'd like to concentrate my efforts on increasing my system's recognition rate. Geometrical and appearance features could be used to describe the action. Finally, I'd like to increase the time efficiency of my system so that it may be used in a variety of applications.

REFERENCES

[1] Mehrabian G. "Nonverbal communication", Aldine, New Brunswick, NJ, USA. 2007.

[2] Ekman P, Friesen WV, O'Sullivan M, Chan AYC, Diacoyanni-Tarlatzis I, Heider KG, Krause R, LeCompte WA, Pitcairn T, Bitti PER. "Universals and cultural differences in facial expressions of emotion", J Pers Soc Psychol 53(4):712–717. 1972.

[3] Ekman P, Friesen W. "The Facial Action Coding System: A Technique for the Measurement of Facial Movement", Consulting Psychologists Press, Santa Clara, CA, USA. 1978.

[4] Bartlett MS, Littlewort G, Fasel I, Movellan JR. "Real time face detection and facial expression recognition: Development and applications to human computer interaction", Proc IEEE Conf Comput Vis Pattern Recog Workshop 5:53–53. 2003.

[5] Whitehill J, Serpell Z, Lin YC, et al . "The faces of engagement: Automatic recognition of student engagement from facial expressions", IEEE Trans Affect Comput 5:86–98. 2014.

[6] Soleymani M, Pantic M. "Emotionally Aware TV", Proc TVUX-2013 Work Explor Enhancing User Exp TV ACM CHI 2013.

[7] Wang Q, Jia K, Liu P. "Design and Implementation of Remote Facial Expression Recognition Surveillance System Based on PCA and KNN Algorithms", Proc - 2015 Int Conf Intell Inf Hiding Multimed Signal Process IIH-MSP 2015, pp 314–317. 2016.

[8] Cockburn J, Bartlett M, Tanaka J, Movellan J, Pierce M, Schultz R. "SmileMaze: a tutoring system in real-time facial expression perception and production in children with autism spectrum disorder", In: Proceedings of the workshop facial bodily expressions control adaptation games. 2008.

[9] Reddy CVR, Reddy US, Kishore KVK. "Facial emotion recognition using NLPCA and SVM", Trait du Signal 36:13–22. 2019.

[10] Mahesh Babu D, Venkata Rami Reddy Ch, Srinivasulu Reddy U. "An automatic driver drowsiness detection system using DWT and RBFNN", Int J Recent Technol Eng 7(5S4):41–44. 2019.

[11] Mayya V, Pai RM, Manohara Pai MM. "Automatic Facial Expression Recognition Using DCNN", Procedia Comput Sci 93:453–461. 2016.

[12] Mahesh Babu D, VenkataRamiReddy Ch, Srinivasulu Reddy U. "An automatic driver drowsiness detection system using DWT and RBFNN", Int J Recent Technol Eng 7(5S4):41–44. 2019.

[13] Gogić I, Manhart M, Pandžić IS, Ahlberg J. "Fast facial expression recognition using local binary features and shallow neural networks", Vis Comput 36:97–112. 2020.

[14] Lee SH, Plataniotis KN, Ro YM. "Intra-Class Variation Reduction Using Training Expression Images for Sparse Representation Based Facial Expression Recognition", In: IEEE Transactions on Affective Computing, vol. 5, pp 340–351. 2014.

[15] Xie S, Hu H. "Facial expression recognition using hierarchical features with deep comprehensive multipatches aggregation convolutional neural networks", IEEE Trans Multimedia 21:211–220. 2019.

[16] Xie S, Hu H, Wu Y. "Deep multi-path convolutional neural network joint with salient region attention for facial expression recognition", Pattern Recognit 92:177–191. 2019.

[17] Ramireddy C V., Kishore KVK. "Facial expression classification using Kernel based PCA with fused DCT and GWT features", 2013 IEEE Int Conf Comput Intell Comput Res IEEE ICCIC, vol. 2013, pp 2–7. 2013.

[18] Pons G, Masi D. "Supervised committee of convolutional neural networks in automated facial expression analysis", IEEE Trans Affect Comput 9:343–350. 2018.

[19] Li Y, Zeng J, Shan S, Chen X. "Occlusion aware facial expression recognition using CNN with attention mechanism", IEEE Trans Image Process 28(5):2439–2450. 2019.

[20] Li Y, Shi H, Chen L, Jiang F. "Convolutional approach also benefits traditional face pattern recognition algorithm", [208!] International Journal of Software Science and Computational Intelligence, vol. 11, pp 1–16. 2019.

[21] Purnama J, Sari R. "Unobtrusive academic emotion recognition based on facial expression using rgb-d camera using adaptive network-based fuzzy inference system (ANFIS)", Int J Softw Sci Comput Intell 11:1–15. 2019.

[22] Alsmirat MA, Al-Alem F, Al-Ayyoub M, Jararweh Y, Gupta B. "Impact of digital fingerprint image quality on the fingerprint recognition accuracy", Multimedia Tools and Applications 78(3): 3649–3688. 2019.

[23] Connie T, Al-Shabi M, Cheah WP, Goh M. "Facial expression recognition using a hybrid CNN–SIFT aggregator", In: Proceedings of the MIWAI, Cham, Switzerland Springer, vol 10607. Pp 139–149. 2017.

[24] Gonzalez-Lozoya S, de la Calleja J, Pellegrin L, Escalante HJ, Medina M, Benitez-Ruiz A. "Recognition of facial expressions based on CNN features", Multimedia Tools Appl 79:13987–14007. 2020.

[25] Ozcan T, Basturk A (2020) Static facial expression recognition using convolutional neural networks based on transfer learning and hyperparameter optimization. Multimedia Tools and Applications 79:26587–26604

[26] Verma, G. and Verma, H., 2020. Hybrid-Deep Learning Model for Emotion Recognition Using Facial Expressions. The Review of Socionetwork Strategies, 14(2), pp.171-180.

AUTHOR PROFILE

Tadese Henok Seifu: Received a B.Sc. degree in Information Science from Mekelle University, Ethiopia in 2017. Afterwards, worked for one-year in Federal TVET Institute as Technical Assistant in Information Communication Technology Department. Now, he is Postgraduate student with a specialization in Software Engineering at Tianjin University of Technology and Education, China (2019-2022). His research interest area is Computer application, Machine learning and intelligent computing.