Psychological Parametric Prediction Through Video Recognition

Sairanjan Dasgupta, Vinayak Mathur, Safa M, Kayalvizhi Jayavel

SRM Institue of Science and Technology, Kattankulathur

Corresponding author: Vinayak Mathur, Email: vm1159@srmist.edu.in

Recognizing human expressions and emotions is one of the most powerful and challenging tasks in social communication. In general, facial expressions are a natural and direct way for humans to express the emotions and intentions. To build a system for acknowledging and recognizing the different emotions present in the emotion spectrum (namely: happiness, sadness, fear, anger, disgust and surprise). This paper goes through the various approaches that can be followed for this topic.

Keywords:Human Facial Expressions, Audio Emotion Detection, Face Expression Recognition, Classification, CNN.

2023. In Saroj Hiranwal & Garima Mathur (eds.), *Artificial Intelligence and Communication Technologies*, 807–813. Computing & Intelligent Systems, SCRS, India. https://doi.org/10.52458/978-81-955020-5-9-75

1. Introduction

With the advancement of technology, the only limits Artificial Intelligence is shackled by are our own imagination and skill. Psychometric analysis, which may be utilised for a variety of objectives, is a prominent area where AI can be used. Our goal is to dip our toes into this sphere and see what AI is actually capable of. Nowadays, it is paramount to teach machines about human emo-tions. This is due to the fact that 'Social Intelligence' requires the advancement of this in its field. It also quickens the progress in the field of HCI (Human-Computer Interaction). Machines in the future must be able to interact with human beings like any other human being would. Only then, the term 'Artifi-cial Intelligence' be completely justified. This can be achieved by extensively learning, imitating and differentiating between the different emotions a human being can go through.

2 Different Approaches Followed for FER

2.1 Neural Networks

The neural network has a hidden layer of neurons. The strategy is predicated on the assumption that for each image, the system has access to a neutral face image.

Face Emotion Recognition using CNN is a two-part CNN that eliminates the image's backdrop before focusing on the extraction of facial feature vectors. The expressional vector is used by the FERC model to detect the six kinds of regular facial expressions.

2.2 Gabor Filter

Gabor Filter is a texture analysis filter. It is prevalent in image processing. It ex-amines whether the image contains any certain frequency content in specific di-rections in a constrained region surrounding the point or region of interest.

Despite the absence of evidence, it's hotly debated whether Gabor filters are similar to the human visual system. They are mainly used in two processes:

- Texture Discrimination.
- Texture Representation.

Gabor filters and Gabor wavelets can be built for a variety of dilations and rota-tions and are considered related.

If expansion was needed, computation of bi-orthogonal wavelets (a very time-consuming process) would be required. This is why expansion is generally not done by Gabor wavelets. As a result, a Gabor filter bank with varied sizes and rotations is commonly built. To create a Gabor space the signals and filters are convolved together. This process is identical to what happens in the main visual cortex.

2.3 HOG Feature

A feature descriptor is used to determine the most essential characteristics in any image.
These feature descriptors are utilised to distinguish between the various photos.

Some different types of feature descriptors are:

- HOG
- SURF
- SIFT

Histogram of Oriented Gradients is a window supported frame feature that uses gradient filters. The edge information of the registered face photographs is used to extract the features.

The structure/shape of the object/image is the key concern for a HOG feature descriptor. It is able to identify edge features as well as edge directions. Calcu-lating the gradient and orientation (in localized regions) is done to recognize edge features and direction**.**

2.4 Support Vector Machines

SVMs are 'Supervised Learning Models', these are used in:

- Regression Analysis
- Classification

SVM produces findings that may be divided into two groups. As a result, it is classified as a non-probabilistic binary model. SVM needs to create a decision boundary that divides an n-dimensional space into specific categories. To create the hyperplane, SVM selects the extreme points. These are known as support vectors.

There are two types of SVMs:

- Linear: Which is used when a dataset can be separated into 2 categories by one straight line.
- Non-Linear: When a dataset can't be divided by a straight line.

There can be multiple decision boundaries. But the one that is the best out of all of them is the hyperplane.

3. Algorithms Used

3.1 Convolutional Neural Network

Face recognition technique based on CNN (Convolutional Neural Network) has been the most widely used approach in the field of face recognition since the in-vention of deep learning. The convolutional and downsampled layers of CNN are built using opencv's convolution and downsampling functions to analyse the images. At the same time, MLP's essential premise is to grasp the whole connec-tion and classification layers, and to do so using Python's theano module. The convolution and sampling layers are integrated into a single layer to simplify the CNN model. Improve the picture recognition rate significantly using the already trained network.

3.2 Haar Cascade

A Haar Cascade's main goal is to recognise objects in photos and videos.

To make a Haar Cascade, 4 steps are needed:

- Haar Features Calculation
- Integral Image Creation
- Adaboost
- Cascading Classifiers Implementation

The sum of the image pixels in the darker part of the image and the sum of the image pixels in the lighter part of the image are calculated. The algorithm uses edge or line detection features.

4. Modules Implemented

4.1 Face Detection and Expression Classification

4.1.1 Datasets used

- RAF-DB (Real World Affective Faces Database) : This contains around 30,000 facial images with various features and emotions that were extensively obtained from the internet.
- FER2013: Facial Emotion Recognition Database - The images in this dataset are grayscale images that have a resolution of 48 x 48 pixels.

The faces in the images have already been adjusted and justified so there aren't any unnecessary portions in the image and all the faces take up a similar amount of area.

4.1.2 Facial Feature Extraction

Locations, regions, and landmarks are located in a 2D or 3D range picture in Facial Feature Extraction. This is then used to get a numerical feature vector at this stage. The most common features that are extracted are:

- Eyebrows
- Nose Tip
- Eyebrows
- Lips

This step is critical since it serves as the foundation for subsequent approaches like facial expression recognition., face tracking, etc. The most important fea-ture to be found is the eye since the other feature locations can be found using it.

4.1.3 Facial Registration

Face Registration is a technique for recognizing and identifying human faces.This, in turn, is utilized for a variety of applications.Facial Localization reveals significant features that should be taken from the faces in the image.These faces are then conformed to match a template image.

4.1.4 Preprocessing

Raw data is transformed into well-formed data sets in the preprocessing step.This is

done to utilize data mining processes.

The most commonly used preprocessing processes are:

- Reducing noise in the image
- Making the image grayscale or binary
- Altering the pixel brightness in the image.
- Geometric Transformation.

Raw data cannot be used consistently since it is incomplete and inefficient most of the time. But if data analytics is to be done on this data, it needs to be adequate. How well a project performs depends on how well the data is format-ted.

4.1.5 Emotion Classification

The aim of this step is to compartmentalize the face on the basis of the six basic human sentiments.

These six emotions are:

- Happiness
- Sadness
- Anger
- Fear
- Surprise
- Disgust

There is a seventh 'neutral' emotion as well which serves as the base for com-paring emotions.

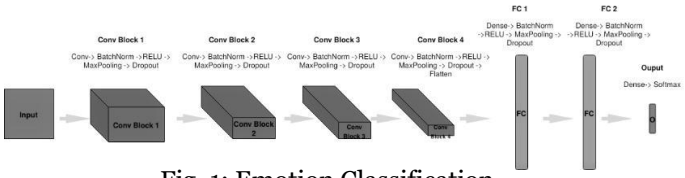

Fig. 1: Emotion Classification

4.2 Audio Detection and Emotion Classification

4.2.1 Datasets used:

(i) **RAVDESS**: Ryerson Audio-Visual Database of Emotional Speech and Song

It is an open source database that contains around 8000 audio files.247 different volunteers have supplied ratings with varying intensity, genuineness, etc. All of these volunteers came from North America.

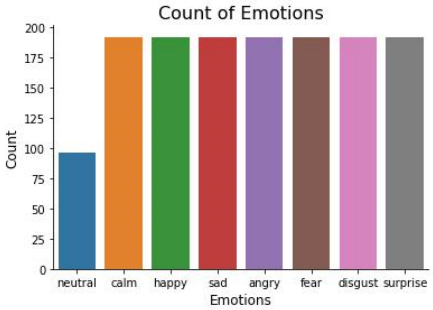

Fig. 2: Graph of Emotions

4.2.2 Audio Feature Extraction and Classification:

For this step, CNNs were used as well.CNNs are well known to work well on images. This is because they can find pat-terns even when they are translation invariant and spatially hierarchical.This basically means that if there are two copies of the same image in the same plane but in different positions, the CNN can identify that both the images are the same image but they aren't in the same location/position.So, we can use features that look like an image and use them in a CNN.

There are two types of images that a CNN will accept:

- A color image with three channels (RGB)
- A grayscale image with one channel. After doing that, a Mel-Spectrogram is needed.

MFCC - Mel-frequency cepstral coefficients. These are the values that make up a Mel-Spectrogram.This results in sculpting the audio features like an image.

5. Results

From the above analysis, we have decided to go with the CNN based implemen-tation for our project.

- That, used in tandem with the various datasets mentioned above, gives us acceptable results.
- The accuracy we got from our project was: 68 percent

6. Conclusion

The emotion recognition system implemented in this research work is a robust model that maps behavioral and psychological characteristics. The psychologi-cal characteristics such as happiness, sadness, anger, fear, surprise are directly correlated to the geometrical structures and positions of the facial features. This topic can be further implemented to design asymmetric cryptosystems that will make systems like passwords and smart cards obsolete.

References

[1] Zhou, H., Meng, D., Zhang, Y., Peng, X., Du, J., Wang, K., Qiao, Y.:Exploring emotion features and fusion strategies for audio-video emotion recognition. 2019 International Conference on Multimodal Interaction (2019). https://doi.org/10.1145/3340555.3355713

[2] Metallinou, A., Lee, S., Narayanan, S.: Audio-visual emotion recognition using gaussian mixture models for face and voice. In: 2008 Tenth IEEE International Symposium on Multimedia, pp. 250–57 (2008). https://doi.org/10.1109/ISM.2008.40

[3] Gunes, H., Schuller, B., Pantic, M., Cowie, R.: Emotion representation, analysis and synthesis in continuous space: A survey. In: 2011 IEEE Inter- national Conference on Automatic Face Gesture Recognition (FG), pp.827–834 (2011). https://doi.org/10.1109/FG.2011.5771357

[4] Zheng, W., Xin, M., Wang, X., Wang, B.: A novel speech emotion recognition method viaincomplete sparse least square regression. IEEE Signal Processing Letters 21(5), 569–572 (2014). https://doi.org/10.1109/LSP. 2014.2308954

[5] Trigeorgis, G., Ringeval, F., Brueckner, R., Marchi, E., Nicolaou, M.A.,Schuller, B., Zafeiriou, S.: Adieu features? end-to-end speech emotion recognition using a deep convolutional recurrent network. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp 5200–5204 (2016). https://doi.org/10.1109/ICASSP.2016. 7472669

[6] Gemmeke, J.F., Ellis, D.P.W., Freedman, D., Jansen, A., Lawrence, W.,Moore, R.C., Plakal, M., Ritter, M.: Audio set: An ontology and human-labeled dataset for audio events. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 776–780 (2017). https://doi.org/10.1109/ICASSP.2017.7952261

[7] Yoon, S., Byun, S., Jung, K.: Multimodal speech emotion recognition using audio and text. In: 2018 IEEE Spoken Language Technology Workshop (SLT), pp. 112–118 (2018). https://doi.org/10.1109/SLT.2018. 8639583 Springer Nature 2021 LATEX template Article Title 13

[8] Pantic, M., Rothkrantz, L.J.M.: Automatic analysis of facial expressions: the state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(12), 1424–1445 (2000). https://doi.org/10.1109/ 34.895976

[9] Zhao, X., Zhang, S.: A review on facial expression recognition: Feature extraction and classification . IETE Technical Review 33, 1–13 (2016).https://doi.org/10.1080/02564602.2015.1117403

[10] Hashim Abdulsalam, W., al-hamdani, d.r., Al Salam, M.: Emotion recognition system based on hybrid techniques. International Journal of Machine Learning and Computing 9, 490–495 (2019).https://doi.org/10. 18178/ijmlc.2019.9.4.831

[11] Burkert, P., Trier, F., Afzal, M.Z., Dengel, A., Liwicki, M.: DeXpression: Deep Convolutional Neural Network for Expression Recognition (2016)

[12] da Silva, E.A.B., Mendon ̧ca, G.V.: 4 - digital image processing. In: CHEN, W.-K. (ed.) The Electrical Engineering Handbook, pp. 891–910. Academic Press, Burlington(2005). https://doi.org/10.1016/B978-012170960-0/50064-5. https://www.sciencedirect.com/science/article/pii/B9780121709600500645

[13] Chollet, F.: Xception: Deep learning with depthwise separable convolutions. CoRR abs/1610.02357 (2016) 1610.02357

[14] Torfi, A., Iranmanesh, S.M., Nasrabadi, N.M., Dawson, J.M.: Coupled 3d convolutional neural networks for audio-visual recognition. CoRR abs/1706.05739 (2017) 1706.05739

[15] Fridlund, A.J.: Evolution and facial action in reflex, social motive, and paralanguage. Biological Psychology 32(1), 3–100 (1991). https://doi. org/10.1016/0301-0511(91)90003-Y

[16] Goodfellow, I.J., Erhan, D., Carrier, P.L., Courville, A., Mirza, M., Ham-ner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.-H., Zhou, Y., Ramaiah, C., Feng, F., Li, R., Wang, X., Athanasakis, D. Shawe-Taylor