International Journal of Research in Engineering, Science and Management
Volume-3, Issue-8, August-2020
journals.resaim.com/ijresm | ISSN (Online): 2581-5792 | RESAIM Publishing

Facial Expression Recognition using Support Vector Machine (SVM) and Convolutional Neural Network (CNN)

Afeefa Muhammed[1*], Ramsi Mol[2], L. Revathy Vijay[3], S. S. Ajith[4], A. R. Shamna[5]
[1,2,3]*Student, Department of Electronics and Communication Engineering, Musaliar College of Engineering, Trivandrum, India*
[4]*Assistant Professor, Department of Electronics and Communication Engineering, Musaliar College of Engineering, Trivandrum, India*
[5]*Associate Professor, Department of Electronics and Communication Engineering, Musaliar College of Engineering, Trivandrum, India*
*Corresponding author: afeefamuhammed98@gmail.com

***Abstract*: Facial expression is an important mode of non-verbal conversation among people. It is really a speedy growing and an evergreen research field in the region of computer vision, artificial intelligence and automation. Facial expressions can be recognized by training images. There are some limitations such as noise in the presently available emotion recognition techniques. This paper proposes a facial expression recognition method based on Support Vector Machine (SVM). It also adopts Convolutional Neural Network (CNN) for image training. SVM follows a technique called kernel trick to transform the data. SVM performs better than other existing techniques, and there by improves the overall performance of facial expression recognition.**

***Keywords*: Biometric markers, Computer vision, Convolutional neural network, Kernel trick, Machine learning, Support vector machine.**

1. Introduction

Humans are emotional creatures. Our emotional state informs how we behave from the most fundamental processes, to complex actions and difficult decisions. Our lives are in many ways guided by our emotions, so knowing more about emotions allows us to know more about human behavior more generally. It's clear that understanding the emotional state of people can be useful for a range of applications from developing a better understanding of human psychology, to investigating behavior for improved user experiences, to developing productive advertising campaigns, and beyond. There are two main categories for the several approaches of facial feature description: geometric feature-based approach and appearance feature-based approach. In the geometric feature-based approach, the geometric relations between facial components are used in the description of a facial image. However, the facial components required for this purpose should be of accurate positions, which become difficult to detect when the appearance of the person changes. Thus, such an approach may experience weaker results under several situations. But appearance feature-based approach explain the overall appearance of the face. There are three main steps included in facial expression recognition. The three steps are following:

1. Face detection: This involves locating face in an image or video footage.

2. Facial landmark detection: This involves extracting information about facial features from detected face. For example, detecting shape of facial components or describing texture of skin.

3. Facial expression and emotional classification: This step involves analyzing the movement of facial features or changes in facial features and classifying information into categories such as facial muscle activation like smile, emotion categories such as happiness or anger, attitude categories like liking or disliking.

Facial expression recognition uses algorithm to detect faces, codes facial expression and recognize emotional states. It does this by analyzing faces in images or videos through cameras embedded in laptops, mobile phones, or computers. It can recognize different emotions on a human face, business images and videos in real time for monitoring video feeds. Facial expression recognition finds application in animated movies, monitor human stress level, extracting expressions of psychiatric patients. It is also applicable in driver's drowsiness detection. Smart cars can alert driver when he is feeling drowsy by first detecting his face and then his eyes. It can also applicable in emotion detection in interview to determine whether the candidate's personality is a good fit for the job. It is also used in the testing for video game. During the testing phase, users are asked to play the game for a given period and their feedback is incorporated to make the final product.

Facial expression recognition technique uses biometric markers to detect emotions in human faces. This technique

automatically detects the six universal expressions. Facial expression conveys non-verbal communication that play an important role in interpersonal relationship. Nowadays different technologies are being used for the recognition of facial expression. The existing technology is the Neighborhood Edge Directional Pattern (NEDP), which is performed in MATLAB. NEDP will not give an output from a live image or video feed. The image quality will be poor and the programming may become slow while performing in MATLAB. So, here we are proposing a new one called Support Vector Machine (SVM) along with Convolutional Neural Network (CNN), which is performed in Python. SVM can recognize the facial expression from a live image and can also work on high dimensional data space. Python has also a user friendly data structure and is faster than other platforms. Thus it contributes a better performance.

2. Proposed Method and Design

The proposed method for facial expression recognition uses the algorithm, Support Vector Machine and neural network, Convolutional Neural Network. The platform used here is Python. Python is a general-purpose, object-oriented, high-level and powerful modern computer programming language. Python uses English keywords frequently where as other languages uses punctuations. It is easy to learn, easy to read, easy to maintain, portable and extendable language. It is processed at run time by the interpreter. You do not need to compile your program before executing it. It can be used to handle big data and perform complex mathematics. The block diagram for proposed method is shown in fig. 1.

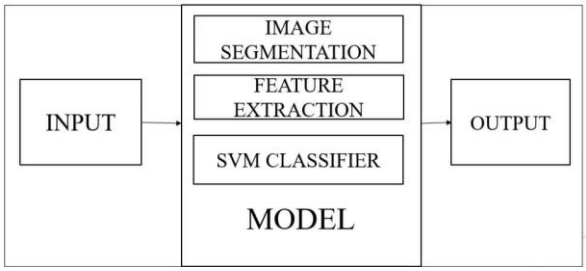

Fig. 1. Block diagram for proposed method

The input is the image. Each image consists of dataset values. The input is obtained by using web cams or it is store it on the OS folder. The image is then sent for segmentation. It involves dividing a visual input into segments to simplify image analysis. Segments represent objects or parts of objects, and comprise sets of pixels. Segmentation is done by using the top-down approach, which is the breaking down of the system into sub-systems. The feature of the segmented image is then extracted. For converting the gradient value to grey scale value, a method called average method is used. The expression for average method is $\frac{(R+G+B)}{3}$ where R, G and B represent pixel components red, green and blue. The extracted feature is then fed to the SVM for classification. The training section is done by the CNN. The image would be trained for more number of times repeatedly. The trained image is then compared with the image we want to test. Then the output is displayed in the OS screen.

A. Support Vector Machine (SVM)

Support Vector Machine (SVM) was first heard in 1992, introduced by Boser, Guyon, and Vapnik in COLT-92. Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression. They belong to a family of generalized linear classifiers. In another terms, Support Vector Machine (SVM) is a classification and regression prediction tool that uses machine learning theory to maximize predictive accuracy while automatically avoiding over-fit to the data. Support Vector machines can be defined as systems which use hypothesis space of a linear functions in a high dimensional feature space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory. Support vector machine was initially popular with the Neural Information Processing System (NIPS) community and now is an active part of the machine learning research around the world. SVM becomes famous when, using pixel maps as input; it gives accuracy comparable to sophisticated neural networks with elaborated features in a handwriting recognition task. It is also being used for many applications, such as hand writing analysis, face analysis and so forth, especially for pattern classification and regression based applications. The foundations of Support Vector Machines (SVM) have been developed by Vapnik. The classification of image can be done in linearly and non-linearly separable data. Classification in linearly and nonlinearly separable data is shown in fig. 2 and fig. 3.

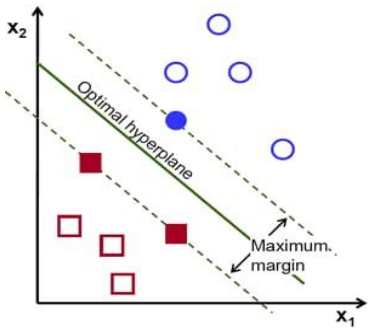

Fig. 2. Classification of linearly separable data

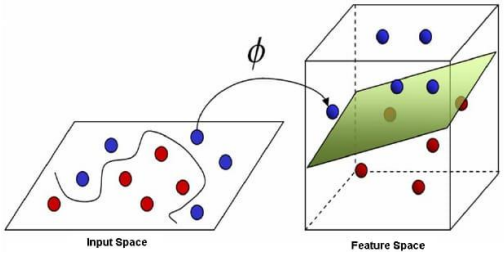

Fig. 3. Classification of non-linearly separable data

Classifying data is a common task in machine learning. Suppose some given data points each belong to one of two classes, and the goal is to decide which class a new data point will be in. In the case of support-vector machines, a data point is viewed as a p-dimensional vector, and we want to know whether we can separate such points with a (p-1)-dimensional hyper-plane. This is called a linear classifier. There are many hyper-planes that might classify the data. One reasonable choice as the best hyper-plane is the one that represents the largest separation, or margin, between the two classes. So we choose the hyper-plane so that the distance from it to the nearest data point on each side is maximized. If such a hyper-plane exists, it is known as the maximum margin hyper-plane and the linear classifier it defines is known as a maximum-margin classifier. However, SVM can be used for classifying a nonlinear dataset. This can be done by projecting the dataset into a higher dimension in which it is linearly separable. In machine learning, a trick known as "kernel trick" is used to learn a linear classifier to classify a non-linear dataset. It transforms the linearly inseparable data into a linearly separable one by projecting it into a higher dimension. A kernel function is applied on each data instance to map the original non-linear data points into some higher dimensional space in which they become linearly separable. Using kernel function, the dot product between two vectors are obtained so that every point is made to a high dimensional data space. Kernel functions are mainly of four types- linear, polynomial, radial basis function (rbf) and sigmoid. Here the kernel function used is the radial basis function. The expression for this function is:

$$e^{\gamma|uv|^2}$$

Where γ is the kernel co-efficient for rbf function, u is the testing vector and v is the support vector.

B. Convolutional Neural Network (CNN)

In neural networks, Convolutional neural network (ConvNets or CNNs) is one of the main categories to do images recognition, images training. Objects detections, recognition faces etc., are some of the areas where CNNs are widely used. CNN were inspired by biological process in that the connectivity pattern between neurons resembles the organization of human visual cortex. Computers sees an input image as array of pixels and it depends on the image resolution. Based on the image resolution, it will see h x w x d (h = Height, w = Width, d = Dimension). The basic block diagram of CNN is shown in fig. 4. The CNN consist of three layers: Convolution layer, Pooling Layer and Fully connected layer.

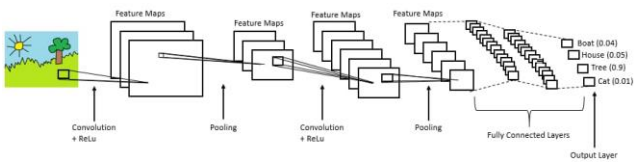

Fig. 4. Block diagram of CNN

1) Convolution Layer

Convolution is the first layer to extract features from an input image. Convolution preserves the relationship between pixels by learning image features using small squares of input data. It is a mathematical operation that takes two inputs such as image matrix and a filter or kernel. The dimensions are as follows:

Image matrix; h×w×d

Filter ; $f_h \times f_w \times d$

Output ; (h-f_h+1)×(w-f_w+1)× 1

This output is referred to as feature map. ReLU here stands for Rectified Linear Unit for a non-linear operation. The output is $f(x) = \max(0,x)$.

2) Pooling Layer

Pooling layers' section would reduce the number of parameters when the images are too large. Spatial pooling also called subsampling or down sampling which reduces the dimensionality of each map but retains important information. Spatial pooling can be of different types:

- Max Pooling
- Average Pooling
- Sum Pooling

Max. pooling takes the largest element from the rectified feature map. Taking the largest element could also take the average pooling. Sum of all elements in the feature map call as sum pooling.

3) Fully Connected Layer

The layer we call as FC layer, we flattened our matrix into vector and feed it into a fully connected layer like a neural network. It acts as human neuron, which interconnects with one another for the transmission of information. All the feature maps from the pooling section are interconnected to provide the trained output. After this, the output layer will recognize the trained images. This trained image is used for comparing with the new image. After comparison the apt expression is recognized.

3. Result and Discussion

The desired facial expressions can be obtained by using this method. Six universal expressions such as happiness, sadness, anger, fear, disgust, and surprise, can be detected here. Some of the expressions are shown in Fig.5. Multiple faces can be detected in the screen at a time. This means multiple expression can be obtained. The solution for this technique is guaranteed to a large extend. This can be applicable for linearly and non-linearly separable data. For high dimensional data spaces and places where calculations become complex, SVM and CNN can be implemented.

The accuracy level of the system is 70- 85%. For obtaining more accurate output, more training of image is required. Hence image training is an important factor in expression recognition. Facial expression recognition also depends on the presence of light. Improper supply of light will also affect its accuracy level. SVM will not be able to classify the image if the choice of kernel function fails. So proper choice of kernel function is

important. In this project we have proposed a combination of two methods: SVM and CNN to extract the facial expressions. By using these techniques, we can detect the real time images and it helps to increases the speed of classification. The classified features have good accuracy displaying the expression and facial action units.

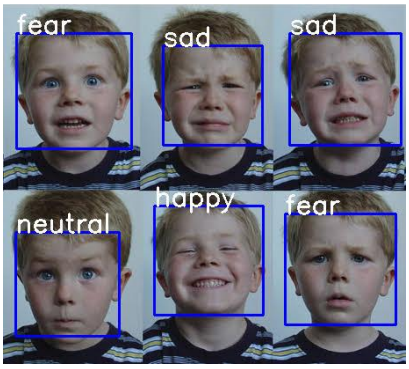

Fig. 5. Detected facial expressions

4. Conclusion

Facial feature tracking and facial actions recognition from image sequence attracted great attention in computer vision field. Computational facial expression analysis is a challenging research topic in computer vision. It is required by many applications such as human computer interaction, computer graphic animation and automatic facial expression recognition. Facial expression recognition or computer-based facial expression recognition system is important because of its ability to mimic human coding skills. Facial expressions and other gestures convey nonverbal communication cues that play an important role in interpersonal relations.

In this paper, facial expressions can be obtained by using two methods. Six universal expressions can be recognized here. We can obtain multiple faces on the screen. We have adopted a mixing of two approaches: SVM and CNN to extract the facial features and expressions. In which SVM is used for the classification of images with kernel trick and CNN is used for the training of images. In future we can obtain more facial expressions by giving more training to the images using CNN.

References

[1] Caifeng Shan, Shaogang Gong, Peter W. Mc Owan, "Facial expression recognition based on Local Binary Patterns: A comprehensive study," Image and Vision Computing, Volume 27, Issue 6, 2009, Pages 803-816.
[2] Vasanth P.C., Nataraj K.R, "Facial Expression Recognition Using SVM Classifier," Indonesian Journal of Electrical Engineering and Informatics, vol. 3, no. 1.
[3] M. Pantic and M. S. Bartlett "Machine analysis of facial expressions", I-Tech Education and Publishing, 2007.
[4] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. OliveiraSantos, "Facial expression recognition with convolutional neural networks: coping with few data and the training sample order," Pattern Recognition, vol. 61, pp. 610–628, 2017.
[5] C. Shan, S. Gong, and P. W. McOwan, "Robust facial expression recognition using local binary patterns," in IEEE International Conference on Image Processing 2005, vol. 2. IEEE, 2005, pp. II– 370.
[6] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multi class support vector machines," IEEE transactions on Neural Networks, vol. 13, no. 2, pp. 415–425, 2002.
[7] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on. IEEE, 2016, pp. 1–10.
[8] R. C. Gonzalez and R. E. Woods, "Digital image processing," Prentice Hall, 2008.
[9] P. Khorrami, T. Paine, and T. Huang, "Do deep neural networks learn facial action units when doing expression recognition?" in Proceedings of the IEEE International Conference on Computer Vision Workshops, 2015, pp. 19–27.
[10] "Understanding of Convolutional Neural Network (CNN) — Deep Learning" [Online]. Available: https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnndeep-learning-99760835f148
[11] C. Cortes and V. Vapnik. Support vector networks. Machine Learning, 20:273 – 297, 1995
[12] Tom Mitchell, Machine Learning, McGraw-Hill Computer science series, 1997.
[13] Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study,".
[14] Burges C., "A tutorial on support vector machines for pattern recognition", in "Data Mining and Knowledge Discovery", Kluwer Academic Publishers, Boston, 1998, (Volume 2).
[15] I. Kotsia and I. Pitas, "Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines," in IEEE Transactions on Image Processing, vol. 16, no. 1, pp. 172-187, Jan. 2007.