# Face Detection and Recognition, Face Emotion Recognition Through NVIDIA Jetson Nano

**Chapter** · September 2020

DOI: 10.1007/978-3-030-58356-9_18



Sergio Márquez-Sánchez
Universidad de Salamanca


Niloufar Shoeibi
Universidad de Salamanca


Juan Manuel Corchado Rodríguez
Universidad de Salamanca - AIR Institute

# Face Detection and Recognition, Face Emotion Recognition Through NVIDIA Jetson Nano

Vishwani Sati[1(✉)], Sergio Márquez Sánchez[2(✉)], Niloufar Shoeibi[2], Ashish Arora[3], and Juan M. Corchado[2,4,5,6]

[1] Amity School of Engineering and Technology, Noida, Uttar Pradesh, India
vishwani.sati@gmail.com

[2] Bisite Research Group, University of Salamanca, Calle Espejo 2, 37007 Salamanca, Spain
{smarquez,niloufar.shoeibi,corchado}@usal.es

[3] Indian Institute of Technology, Dharwad, Dharwad, India
ashisharora.010@gmail.com

[4] Air Institute, IoT Digital Innovation Hub (Spain), 37188 Salamanca, Spain

[5] Department of Electronics, Information and Communication, Faculty of Engineering, Osaka Institute of Technology, Osaka 535-8585, Japan

[6] Pusat Komputeran dan Informatik, Universiti Malaysia Kelantan, Karung Berkunci 36, Pengkaan Chepa, 16100 Kota Bharu, Kelantan, Malaysia

**Abstract.** This paper focuses on implementing face detection, face recognition and face emotion recognition through NVIDIA's state-of-the-art Jetson Nano. Face detection is implemented using OpenCV's deep learning-based DNN face detector, supported by a ResNet architecture, for achieving better accuracy than the previously developed models. The result computed by framework libraries of OpenCV, with the support of the above-mentioned hardware, displayed reliable accuracy even with the change in lighting and angle. For face recognition, the approach of deep metric learning using OpenCV, supported by a ResNet-34 architecture, is used. Face emotion recognition is achieved by developing a system in which the areas of eyes and mouth are used to convey the analysis of the information into a merged new image, classifying the image into displaying any of the seven basic facial emotions. A powerful and a low-power platform, Jetson Nano carried out intensive computations of algorithms easily, contributing in high video processing frame.

**Keywords:** Face recognition · Emotion detection · OpenCV · Deep Neural Network · NVIDIA Jetson Nano

## 1 Introduction

As the data is increasing exponentially, supported by the steady doubling rate of computing power every year, computer vision has become a popular field of research. With researchers highly intrigued about finding insights into how our


P. Novais et al. (Eds.): ISAmI 2020, AISC 1239, pp. 177–185, 2021.
https://doi.org/10.1007/978-3-030-58356-9_18

brain works, computer vision has not limited itself into being only a research area of computer science, but also the object of neuro-scientific and psychological studies. Face detection and recognition, along with the analysis of facial expressions, is currently an active research area in the community of computer vision.

Face Detection is a computer technology which, given a digital image or a video, detects facial features and determines the locations and sizes of human faces, by ignoring anything else, such as trees, buildings or bodies present in the image or video. This localization and detection of human face is a prerequisite for face recognition and/or analysis of facial expressions, used in applications such as video surveillance, image database management and human computer interface.

Face Recognition, introduced by Woodrow Wilson Bledsoe in the 1960s, is being constantly improved and optimized ever since then, becoming gradually mature, and the technology being more and more widely used in human daily life. Then used to confirm the identity of Osama Bin Laden after he was killed in a U.S. raid, the face recognition system is now increasingly used in the smart phones for user authentication and device security, and for forensics by military and law enforcement professionals. Generally, the face recognition process involves two steps: First, the photo is searched to find a face i.e., face detection, by processing the image to crop and extract the person's face for recognition. Second, the detected face is then compared to a stored database of known faces i.e., face recognition, contributing in the identification or verification of one or more persons in a given still or video images of a scene.

A machine can detect and recognize a person's face using a regular web camera. However, factors like viewing a person from an angle, lack of proper lighting or brightness of an image, a blurry picture or a contrast in shadows can significantly increase the difficulty in detecting a face. Since the 1990s, face recognition has been a prominent area of research however, it's still less reliable than face detection and far away from being considered a reliable method of user authentication.

Emotion recognition is a technology that has been gaining lot of attention over the past decades with the development of techniques of artificial intelligence. This can be achieved by inspecting the body posture, voice tone or facial expressions. In this paper, we focus on recognising the emotions using facial expressions. Inferring the facial emotions of other people helps in human communication by understanding the intention of others. Facial Emotion Recognition, being a thriving area of research, has applications in computer animations, human-machine interaction, and in various educational processes – understanding the inner state of mind of the learner.

NVIDIA Jetson Nano, ideally suited as an IoT edge device because of its small size and connectivity options, is used for machine learning inferencing. Considered to be a powerful and a low-power platform, Nano has applications in home robots, intelligent gateways. Delivering 472 GFLOPS of compute performance by utilizing power of only 5 W, Jetson Nano supports high performance

ML acceleration. Researchers have implemented algorithms on NVIDIA Jetson Nano for solving problems concerned with autonomous driving and traffic surveillance, medical and farming, drone navigation.

## 2 Related Work

With technologies of face detection and face recognition being widely used, interest in them date's way back to 1960s when Bledsoe, Chan, and Bisson developed the first face recognition algorithm [1–4]. Most of the resources available for implementing the algorithms to achieve the desired goal of recognition are for Neural Networks, while Eigenfaces work better. Only a few resources like recognition from video and other techniques at the Face Recognition Homepage [5], 3D Face Recognition Wikipedia page [6] and Active Appearance Models page [7] provide explanation to achieve the goal of recognition better than the eigenfaces. However, many other techniques are mentioned in the recent computer vision research papers from CVPR. Computer vision or machine vision conferences, such as CVPR10 and CVPR09 [8], discussed advancements in these techniques, which give slightly better accuracy.

The issue of facial emotion recognition has been an important research area and is being inspected and analysed on various other research areas [9]. Conventional ways of determining the facial emotions have been introduced in some review papers [10,11]. Difference between conventional techniques and deep learning-based approaches for facial emotion recognition has been introduced by Ghayoumi [12] in a review paper. [13] discusses facial emotion recognition in detail. To classify emotions, algorithms like KNN, Random Forest are applied in [14]. The possibility of using deep learning for emotion detection was inferred from the high accuracy rate obtained by using filter banks and Deep CNN [15] to identify emotions from facial images. For recognizing facial emotions, different databases were studied in [16]. Also, Hidden Markov Models and Deep Belief Networks with UAR of approx. 53% have been used to recognize emotions from facial expressions in [17].

With the significant research being done in the fields of IoT, Machine Learning, NVIDIA's Jetson Nano, the latest addition to the Jetson family of embedded computing boards, is being used as an edge computing platform for ML inferencing. Vittorio Mazzia et al. [18] have worked on real-time detection of apples to estimate the apple yields and therefore, manage the apple supplies. Researchers had also worked previously on proposing machine vision system for yield estimation however, those algorithms had high computation power, utilized intensive hardware setup, in addition to which the weight and power constraints made these algorithms unsuitable for real-time apple detection. For machine learning algorithms, Mazzia et al. used Jetson Nano that contributed in accelerating complex machine learning tasks. The light weight, low power consumption and form factor significantly made the goal of yield estimation plausible. Siddhartha S. Srinivasa et al. [19] have devised MuSHR, a considerably economic robotic race car, an open source platform to advance research and education in the field of robotics. The hardware architecture of the robotic car comprises of Nvidia's Jetson Nano, on which the computations are performed. Srinivasa et al. mention that the ability of Nano to be loaded with the desired operating system and program through an SD card has been primary reason of its inclusion in the hardware architecture of MuSHR.

# 3 Methodology

## 3.1 Face Detection

OpenCV's Haar Cascades is popularly used to detect faces in images or videos. However, Haar Cascades serve a disadvantage of not being able to detect the faces that are not at a straight angle. In this paper, we have used OpenCV's DNN Face Detector i.e., a deep learning-based face detector. While MobileNet base network is used by other OpenCV SSDs, DNN Face Detector utilizes ResNet as the base network and the Single Shot Detector (SSD) framework, which enables it to detect faces at angles other than straight angles as well. The algorithm of detecting faces using deep learning OpenCV Face Detectors runs locally and is not cloud based. The version 3.3 of OpenCV has a highly improved module of Deep Neural Networks (DNN). Frameworks supported by this module are Caffe, Tensor Flow and Torch/PyTorch. We have used the Caffe-based face detector in this paper, which requires two set of files i.e., the protxt file(s), used for defining the model architecture, and the caffe model file, containing the weights for the actual layers. The process of face detection involves importing the necessary packages. Additional packages like Video Stream, imutils, and time also have to be imported, only while detecting faces in videos, not in images. Webcam's video feed has been used for face detection in videos. The model is loaded and the video stream is initialized to allow the camera to warm up. To obtain the face detections, the frames from the video stream are looped over in order to pass the blob through the DNN. This enables us to compare the detections to the confidence threshold for the face boxes and the confidence values to be drawn on the screen. After the OpenCV face detections are drawn, the frame is displayed on the screen until a key is pressed to break out of the loop and clean-up is performed.

## 3.2 Face Recognition

Face Recognition is usually performed using deep learning in which a network is typically trained to accept a single input image and then output a classification for that particular image. However, we have used the concept of deep metric learning using OpenCV in this paper, which is slightly different from the deep learning approach. Also, the architecture of the network used is based on ResNet-34. In the deep metric learning approach for face recognition, a real-valued feature vector is obtained as an output instead of a single label output. A list of 128 real-valued numbers is the output feature vector used for quantifying the face in the dlib facial recognition network. The network is trained using triplets. The process of training the network using triplets can be explained step by step as: First, three images are fed to the network as input where two images are of the same person and one is of a different person. The faces in the input images are quantified by the construction of 128-d embedding for each by neural net. While comparing the results, the weights of the neural net are slightly tweaked so that the measurements of the two images of the same person are closer and that of the single image of the different person are further away.

### 3.3 Face Emotion Recognition

In this paper, we have performed face emotion recognition using two major modules: Facial Image Treatment and a propagation algorithm of ANN to recognize the facial expressions. The process can be explained as: First, a new image is provided as an input. Second, it is passed through a series of phases. While doing so, it gets turned into a new merged image which is used for analysis in the ANN. The training set consists of images with seven different face emotions: Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral, with which the ANN has been previously trained. The emotional state of the face is reported by the system once the group to which the image belongs is detected. The human face is constantly analysed by the system and the information regarding the emotional states is extracted using the inputs of the eyes and mouth zone. The extractions are then merged into a single new image, which is resized using the method of Nearest Neighbour Interpolation. The ANN is then provided with the current input data and a back-propagation algorithm, with a feed-forward architecture is used to recognize the facial expressions.

In this paper, face detection, face recognition, and face emotion recognition are implemented using NVIDIA Jetson Nano as the main core as it has high video processing frame rate and also an internet connection is not required. Nano is booted with the Operating System of Ubuntu with libraries like OpenCV, NumPy and Keras installed in it.

## 4 Results

### 4.1 Face Detection

The OpenCV's Deep Learning Face Detector shows better performance in terms of accuracy than the Haar Cascades. Figure 1(a) displays the detected faces with the accuracy of 99.96% for the face closer to the webcam and 99.31% for the detected face a bit farther away. Also, with a powerful board like Nano, intensive computation of DNN Face Detector can be carried out easily. Figure 1(b) displays the robustness of the face detection software as it gives an accuracy of 89.99% even in condition where the lighting is dim. The face is detected with an accuracy of 72.87% in Fig. 1(c), where the face is not at a straight angle. This wouldn't have worked well with the Haar Cascades. Not quite faster, but a robust face detection software is built, with a better accuracy than the Haar Cascades, using OpenCV's deep learning-based face detector.

### 4.2 Face Recognition

The performance of the recognition process improves by adding colour processing, considering application of techniques of edge detection. Also, increasing the number of input images, with images from different angles, different conditions of lighting contributes significantly in the improvement of the accuracy of the recognition procedure. Careful alignment of the pictures and choosing low resolution images over high resolution images give better recognition results. We used the deep learning-based facial embeddings, capable of being executed in real time, which recognized the face with an accuracy of 79.85% (Fig. 2 a) and 86.41% (Fig. 2 b) in different scenarios, different backgrounds. However, it is easier to perform the process of face recognition in real-time because of the same camera, background, expression, lighting or direction of view being used as compared to recognition from a different direction, time or room.

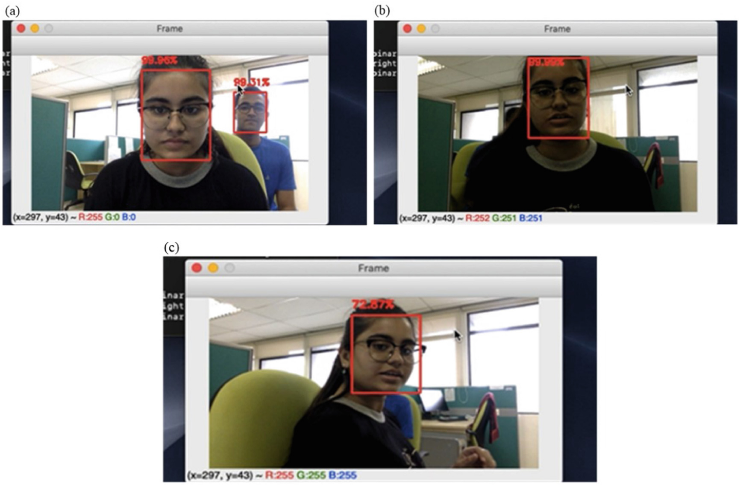


**Fig. 1.** (a) Detected face with the accuracy of 99.96%, (b) Face detection with accuracy of 89.99%, and (c) Face is not at a straight angle.


**Fig. 2.** (a) Face recognition with accuracy of 79.85% (b) Face recognition with 86.41% accuracy.

### 4.3 Face Emotion Recognition

The emotions in static face expressions can be recognized by the methodology adapted in this paper. The images used for training as well as testing purposes were fed as an input to the ANN. Before the feeding of the images to the ANN, a pre-processing step was followed in which the images were resized and merged. A union of the eyes and mouth zone was obtained, as they can be best used for predictions of emotions, reason being the most visual indication of emotions is visible in these zones. Figure 3a and Fig. 3b display the results of face emotion detection by the ANN.

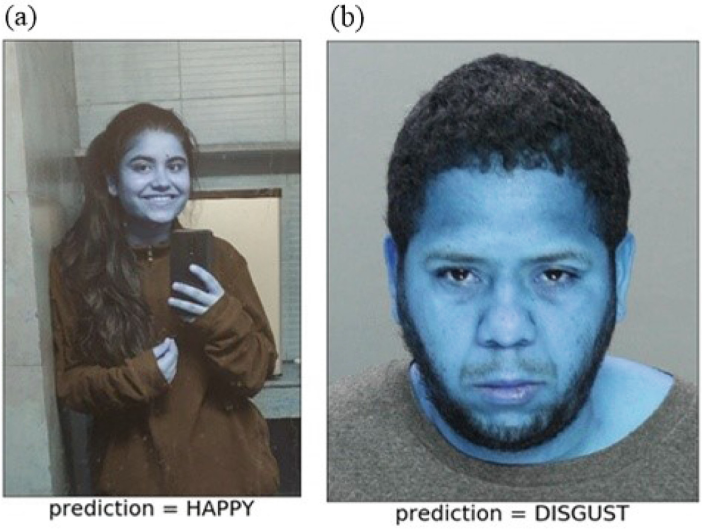


**Fig. 3.** (a and b) Face emotion detection by the ANN

Strengths and weaknesses of the system of face emotion detection is shown in Table 1. Considering the diagonal elements, with 83.3% as the performance of the classifier, all emotions can be classified with an accuracy of more than 75%. Breaking down the details, among the 12 images of anger, the system was successful in classifying 9, while the other 3 were classified as sadness. The perceived reason being that in both the cases of anger and sadness, lips pressed tightly serve as the common denominator. 12 images had the expression of disgust, among which 10 were recognized successfully and the remaining 2 were recognized as sadness. Similarity of the shape of the mouth in both the emotions produced the confusion. With 12 images of fear fed to the system, 9 were successfully classified, remaining 3 were attributed to surprise, as confusion is caused because of the closeness of eyebrows in both the cases. 10 out of 12 images of happiness were classified correctly, rest 2 were recognised as anger, because both the expressions show some teeth areas. Similarly, 10 out of 12 images with the expression of sadness were classified accurately by the system however, 2 were misclassified as disgust. The system could accurately recognize the images with the expression of surprise and neutral expressions.

**Table 1.** Confusion matrix of the system of face emotion detection

| Variables | Anger | Disgust | Fear | Happy | Sad | Surprise | Neutral |
|---|---|---|---|---|---|---|---|
| Anger | 75.00 | 0.00 | 0.00 | 16.67 | 0.00 | 0.00 | 0.00 |
| Disgust | 0.00 | 83.33 | 83.33 | 0.00 | 16.67 | 0.00 | 0.00 |
| Fear | 0.00 | 0.00 | 75.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Happy | 0.00 | 0.00 | 0.00 | 83.33 | 0.00 | 0.00 | 0.00 |
| Sad | 25.00 | 16.67 | 0.00 | 0.00 | 83.33 | 0.00 | 0.00 |
| Surprise | 0.00 | 0.00 | 16.67 | 0.00 | 0.00 | 100.00 | 0.00 |
| Neutral | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 100.00 |

## 5 Conclusion

Face detection and recognition, useful for constructing numerous industrial and commercial applications, is a challenge for many researchers. For making the methodologies more efficient and improving the results, improvement of small features can be done. As technology is advancing, more advanced features can be added to the system, which might be helpful in increasing the accuracy. Also, face emotion recognition is conducted using ANN. Using the same architecture, real time face emotion recognition can be developed, while also increasing the reliability and possibilities.

**Acknowledgments.** This work was supported by the Spanish Junta de Castilla y León, Consejería de empleo. Project: UPPER, aUgmented reality and smart personal protective equipment (PPE) for intelligent pRevention of occupational hazards and accessibility INVESTUN/18/SA/0001.

## References

1. Bledsoe, W.W.: The model method in facial recognition, vol. 15, no. 47, p. 2. Panoramic Research Inc., Palo Alto, CA, Report PR1 (1966)
2. Chan, H., Bledsoe, W.W.: A man-machine facial recognition system: some preliminary results. Panoramic Research Inc., Palo Alto, CA, USA (1965)
3. Bledsoe, W.W.: Some results on multicategory pattern recognition. J. ACM (JACM) **13**(2), 304–316 (1966)
4. Bledsoe, W.W.: Semiautomatic Facial Recognition. Stanford Research Institute, Menlo Park, CA, USA (1968)
5. Face Recognition. http://www.face-rec.org/algorithms/
6. Wikipedia, Three-Dimensional Face Recognition. http://en.wikipedia.org/wiki/Threedimensional_face_recognition

7. Wikipedia, Active Appearance Model. http://en.wikipedia.org/wiki/Active_appearance_model
8. Computer Vision Papers. http://www.cvpapers.com/
9. Swain, M., Routray, A., Kabisatpathy, P.: Databases, features and classifiers for speech emotion recognition: a review. Int. J. Speech Technol. **21**(1), 93–120 (2018)
10. Kołakowska, A.: A review of emotion recognition methods based on keystroke dynamics and mouse movements. In: 2013 6th International Conference on Human System Interactions (HSI), pp. 548–555. IEEE, June 2013
11. Ko, B.C.: A brief review of facial emotion recognition based on visual information. Sensors **18**(2), 401 (2018)
12. Ghayoumi, M.: A quick review of deep learning in facial expression. J. Commun. Comput. **14**(1), 34–8 (2017)
13. Graves, A., Mohamed, A.R., Hinton, G.: Speech recognition with deep recurrent neural networks. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 6645–6649. IEEE, May 2013
14. Mustafa, M.B., Yusoof, M.A., Don, Z.M., Malekzadeh, M.: Speech emotion recognition research: an analysis of research focus. Int. J. Speech Technol. **21**(1), 137–156 (2018)
15. Huang, K.Y., Wu, C.H., Yang, T.H., Su, M.H., Chou, J.H.: Speech emotion recognition using autoencoder bottleneck features and LSTM. In: 2016 International Conference on Orange Technologies (ICOT), pp. 1–4. IEEE, December 2016
16. Le, D., Provost, E.M.: Emotion recognition from spontaneous speech using hidden Markov models with deep belief networks. In: 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 216–221. IEEE, December 2013
17. Harár, P., Burget, R., Dutta, M.K.: Speech emotion recognition with deep learning. In: 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN), pp. 137–140. IEEE, February 2017
18. Mazzia, V., Khaliq, A., Salvetti, F., Chiaberge, M.: Real-time apple detection system using embedded systems with hardware accelerators: an edge AI application. IEEE Access **8**, 9102–9114 (2020)
19. Srinivasa, S.S., Lancaster, P., Michalove, J., Schmittle, M., Rockett, C.S.M., Smith, J. R., Choudhury, S., Mavrogiannis, C., Sadeghi, F.: MuSHR: A Low-Cost, Open-Source Robotic Racecar for Education and Research. arXiv preprint arXiv:1908.08031 (2019)