2021 6th International Conference for Convergence in Technology (I2CT)
Pune, India. Apr 02-04, 2021

A Report on Automatic Face Recognition: Traditional to Modern Deep Learning Techniques

Radha Guha, Ph.D.
CSE Dept.
SRM University AP,
Andhra Pradesh
radha.g@srmap.edu.in

***Abstract*— In the era of smart digital transformation, automatic face recognition is the way of identification and verification of a person in many applications of security and authentication. If the person's 2D still image or 3D video frame is taken under controlled lighting and frontal face poses, then today automatic face recognition is a solved problem with more than 98% accuracy. Under ideal condition of images, automatic face recognition outperforms manual face recognition rate. But the problem of uncontrolled illumination, occlusion, tilting of faces, different expressions of faces, use of accessories and hair color, growth of facial hair, aging effect of a person and low-resolution images makes automatic face recognition underperform; it gets defeated by human. Unless this problem is solved on real time, automatic face recognition system cannot be trusted for crucial security applications like e-passport, fraud detection, counter terrorism and mug-shot verification etc. Extensive research is underway to improve this technology. Traditionally Harr cascade classifier methods, histogram of oriented gradients (HoG), principal component analysis (PCA), Eigen-faces, and support vector machine (SVM) classifier are used in face recognition. Now modern deep learning innovation has superseded those traditional machine learning techniques with more computing power in GPUs and TPUs to tackle all variations in input image quality and ever-growing face database. The goal of this paper is to report the journey of face recognition system from traditional to modern techniques to increase precision and recall of the system.**

Keywords*— *Face Recognition, Machine Learning, HoG, PCA, SVM, Deep Learning, CNN, FaceNet

I. INTRODUCTION: FACE RECOGNITION SYSTEM

Biometric [1] is a unique characteristic of a human viz. fingerprint, retina, face, voice etc. that can be measured and used for person identification. Now a day use of biometrics for person identification has become a part of our daily lives in applications such as office attendance taking, access to computers, access to offices and apartment complexes, access to bank ATM and voter verification etc. Over other biometric techniques for identification like fingerprint recognition, retina scan recognition, voice recognition; face recognition has advantage as it is non-intrusive and does not need target's active cooperation with the sensor instruments. Face recognition is also more challenging than other biometric modalities as input data is collected in unconstrained environment. Just being within the range of a camera a person's image can be captured where the face may not be clearly visible. This is termed as faces in-the-wild.

Automatic face recognition [1], [2], [3], [4], [5], [13] is a non-trivial task of computer vision, where zero or multiple faces must be localized in the entire image first and then identified by matching them with the stored labelled photographs of the persons in the database. Images acquired by camera can vary in quality due to bad lighting condition or for different poses or different facial expressions of the target. It may happen that currently the target has aged than his stored picture in the database. People are seen wearing face masks occluding most of the facial attributes during COVID-19 pandemic situation in 2020-21. It also happens that the target intentionally wants to fool the system by wearing a mask.

Face detection and face recognition is a subfield of computer vision and pattern recognition problem. Face recognition research in the last three decades have gained importance for security and law enforcement applications and has made remarkable progress. Still images taken by camera under controlled lighting condition has all facial features visible and is an easy problem to solve by computer. Modern day cameras from early 2000 put a bounding box surrounding a face to tell that it is rightly focusing the lens on the face. Now a day social networking site FaceBook can detect as well as recognize a face automatically. It will tag friends' faces in an uploaded photograph which was a manual procedure few years ago.

But for surveillance video images we cannot expect perfectly centred faces. Surveillance images are often blurry and under high variability conditions [16] and they create difficulty for automatic face recognition by computers. Whereas human brain is very good at mapping images even when taken under poor lighting and different face expression condition to the ideal face image they have seen before. Another challenge of automatic face recognition is that face recognition must be computed on real time on very large and rapidly growing face database of today. This most important application of modern age must attain human level of perfection in detecting, recognizing and verifying faces in all adverse conditions.

We use the term face recognition and face verification as synonyms. But they are slightly different. Face recognition answers the question 'who is this person?' Face recognition is a difficult multi-classification problem in machine learning. For face recognition there are too many classes as everyone is a separate class and often there are very few samples of each class. Face recognition needs an individual photograph to be compared with all other labelled photographs in the database and generate a ranked list of matches. Face verification answers the question 'is this person same as what he claims he is?' It is a binary classification problem. Face verification needs an individual photograph to be compared with another labelled photograph only. The performance metrics for face recognition is precision and recall. Whereas it is false positive rate (FPR)

2021 6th International Conference for Convergence in Technology (I2CT) | 978-1-7281-8876-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/I2CT51068.2021.9418068

Authorized licensed use limited to: Deakin University. Downloaded on July 26,2021 at 05:43:21 UTC from IEEE Xplore. Restrictions apply.

and false negative rate (FNR) for face verification. The formula for precision, recall, FPR and FNR are as follows: Precision = TP/(TP+FP), Recall =TP/(TP+FN), FPR = FP/(TN+FP), and FNR =FN/(TP+FN). Here TP is true positive rate, FP is false positive rate, TN is true negative rate and FN is false negative rate.

Another issue with image processing in general is that image transformation from spatial domain to frequency domain is a standard practice for image compression. Image compression [3] reduces storage requirement and reduces communication bandwidth requirement. In JPEG image standard the steps followed for image compressions are discrete cosine transform or discrete wavelet transform (DCT/DWT), quantization and entropy encoding. DCT has very good energy compaction capability in its low frequency terms and so higher frequency terms can be discarded for data compression. Similarly, for image decompression the steps are reversed as entropy decoding, inverse quantization and inverse DCT.

At the receiver end when an image is decompressed for display or for other image processing tasks, the decompressed image loses some information depending on how much it was compressed before. Human visual system cannot perceive this loss of information unless the compression ratio is too high i.e., more than 80%. If face recognition algorithm starts with decompressed image input, it must find out whether the compression ratio has any detrimental effect on face recognition accuracy and how much compression can be tolerated. Fortunately, literature review reports that image compression has no significant effect on face recognition accuracy unless the compression ratio is too high i.e., more than 90%.

In other studies, it is reported that face recognition can be performed in compressed domain as well to save computation time of image decompression. DCT requires $O(N)^2$ operations, so is IDCT algorithm, where N is the number of pixels in the original image. So, $O(N)^2$computation can be saved if image decompression is not needed.

For doing face recognition algorithm we need lots of labelled face data. There are many open-source face databases [6], [7] available online for face recognition research. Some of the names are facial recognition technology (FERET), labelled faces in the wild (LFW), Google, FaceBook, MegaFace, YouTubeFace (YTF), and MS-Celeb-10M datasets etc. FERET has 14,126 images of 1199 people. LFW has 13K images of 5K people. MS-Celeb has 10M images of 1M celebrities in the world collected from the web. These are all labelled datasets with face identification. For large datasets like these some incorrect labelling may be there, and it is termed as label noise. Consequence of label noise [16] is less accuracy and more time to train a model requiring larger dataset. IMDB-FACE is a clean dataset with 1.7M images of 59K celebrities created from movie screenshot of IMDB movie database. These databases grow over time with diversity and scale. Anyone can use this large dataset to develop new techniques to improve face detection and recognition rate. As deep learning techniques are data hungry, they can utilize large dataset like MegaFace and IMDB-Face.

A brief introduction of input image condition, face recognition system objectives and available face databases is given in this section. In the next sections this paper explores various steps of face recognition with traditional techniques first and then with modern deep learning approaches which attains better accuracy. There are myriads of research paper in this field. The goal of this paper is to capture the general transition of the trend from traditional machine learning techniques to modern deep learning techniques by reviewing few landmark methods. Section II describes a few traditional machine learning algorithms and their short comings. Section III describes a few breakthrough deep learning models which attains better accuracy and what remains unsolved. Section IV concludes the paper.

II. Traditional Methods

The steps for face recognition are face detection, face alignment, face representation and face matching as shown in Figure 1. Face detection means whether there is a face in the input image and if yes then localizing the face with a bounding box co-ordinates. Face detection system outputs zero, one or multiple bounding boxes surrounding the faces in a photograph. The detected faces are input to the face recognition module. Face recognition means what is the name of the person in the input image. Only the cropped faces inside the bounding boxes are the input to the face recognition system. If the face is not frontal but oriented in a different direction, then face recognition will be a problem. Human can still recognize faces with different orientation. But picture taken from different angles may look very different for a computer to recognize the face. So before feeding to the face recognition module, data is pre-processed in second step called face alignment. To solve the alignment problem [11], several landmarks like centre of eyes, nose, mouth (usually 64) on the face is first determined. Then the eyes and nose are repositioned in the centre by affine image transformation of rotating, shifting, scaling and shearing. After face alignment, in the third step a face is represented with its reduced feature sets [2], [12] derived by algorithms like PCA, IDA and LDA etc. Feature extraction algorithms retain only the most important features of the image that can be later used for image classification or identification. In fourth step for face recognition a supervised classification algorithm like *K*-nearest neighbour or SVM will be used for face matching.

Fig. 1. Face Recognition Steps

Authorized licensed use limited to: Deakin University. Downloaded on July 26,2021 at 05:43:21 UTC from IEEE Xplore. Restrictions apply.

For object detection, shape of the object and edges of the object are most important features. As color information is superfluous, for face detection a color image is first transformed into a simplified grey scale image. Also, the input image size is changed to a fixed lower dimension image size like 64x128 pixels etc., that brings down the computation requirement. In Figure 2, six images of face and non-face are processed for face detection. In Figure 2 top row shows original images, middle row shows edge images processed by Laplacian of Gaussian (LoG) filter and bottom row shows Histogram of Gradients (HoG) detector images. Then these images are matched with face template of varying size to detect faces as shown in Figure 3. Out of six original images in Figure 2, faces in three images are detected as shown in Figure 3.

There are many face detection algorithms [8], [9], [10] in the literature viz. Laplacian of Gaussian (LoG), Viola-Jones detector, Histogram of Gradient (HoG), scale invariant feature transform (SIFT), deformable parts model (DPM) and deep learning with pyramids etc. In 2001 Viola and Jones [9] devised edge and gradient based algorithm for rapid object detection using cascades of weak to more complex features. This approach quickly discards the background and concentrate on object like region, saving computation time. This method achieved faster computation time (approximately 15 times) and better accuracy rate than previous state-of-art techniques in face detection. One drawback of Viola-Jones detector is it needs multi-scale parameter tuning which also varies from image to image. If the parameters per image are not tuned properly it will falsely detect a face or miss a face all together.

In 2005, Dalal and Triggs [10] devised a method named histogram of oriented gradients (HoG) which works better in solving uncontrolled illumination problem in human detection. HoG considers the direction of illumination changes than the actual pixel values and has higher face detection rate. HoG was experimented on MIT pedestrian dataset where it performed with 100% accuracy. So, they created more challenging INRIA dataset with 1805 images of size 64x128 pixels. In these images people are usually standing but with wide range of pose and background variations and poor illumination. HoG outperforms Viola-Jones feature descriptors in detecting faces by reducing false positive rate (FPR) and false negative rate (FNR).

The detail of HOG algorithm [10] is as follows. First it determines gradient of illumination changes of each pixel, i.e., whether it is toward left, right, up, down, up-left, up-right, down-left, down-right etc. Gradient magnitude is given by using this formula $G = \sqrt{G_x^2 + G_y^2}$ and the orientation is given by the formula $\tan(\Phi) = \frac{G_y}{G_x}$. Then gradient orientation is approximated in a block size of 16x16, to the direction that gets majority vote in that block. After this operation, an image will look like in Figure 2 bottom row. Figure 2 captures the basic structure of a face if it is there in the image. Then the entire image is scanned to see if any section of the image matches with a standard face template in HoG. This standard template is a trained HoG output with lots of input face images. HoG descriptors are invariant to small displacement, rotation and illumination changes in the images.

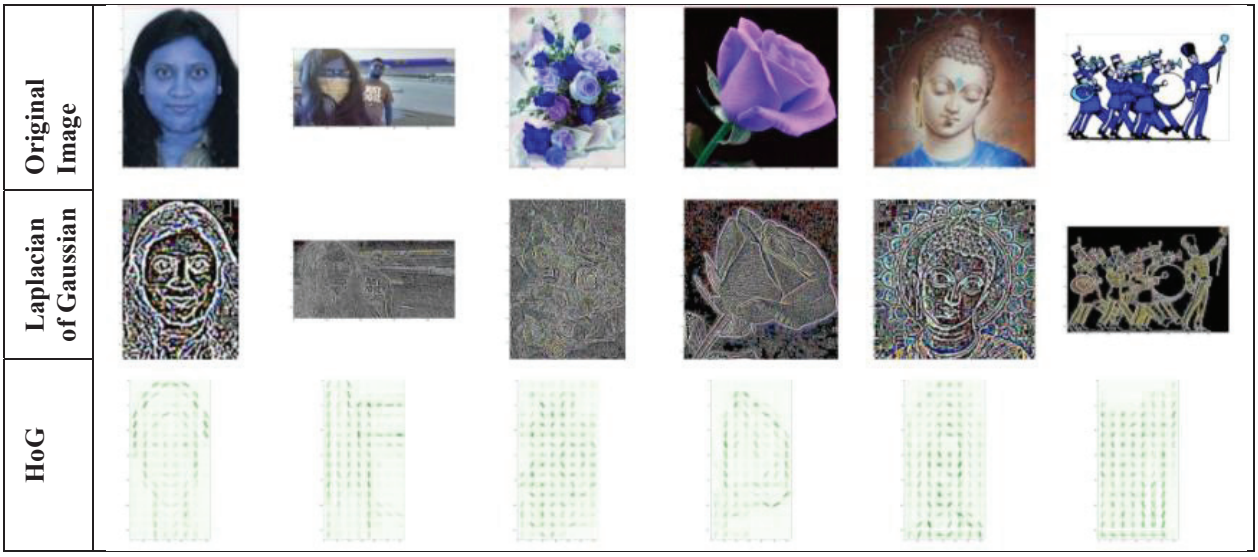

Fig. 2. Top: Original Image, Middle: Laplacian of Gaussian (LoG), Bottom: Histogram of Oriented Gradient (HOG) Techniques

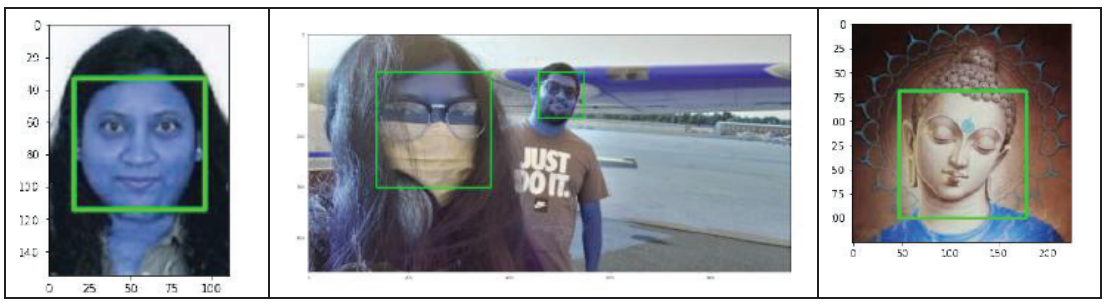

Fig. 3. Detected Faces of Different Sizes from HoG Detector

Authorized licensed use limited to: Deakin University. Downloaded on July 26,2021 at 05:43:21 UTC from IEEE Xplore. Restrictions apply.

In general, after an input image is transformed with any of the above-mentioned feature descriptors like Viola-Jones or Dalal-Triggs, a sliding window of multiple scale will scan the entire image for face detection. In this step there can be multiple overlapping bounding boxes surrounding a face be created, which is processed again for non-maximum suppression. The detected faces inside the bounding boxes are shown in Figure 3. OpenCV provides both Viola-Jones and HoG-detector functions and it takes very few lines of Python code to implement them.

The methods of face recognition can be broadly divided into features extracted from image based and whole image based (holistic). In feature-based face recognition it can be geometric features such as size of head or size of eyes, nose and mouth and their relative distances. In feature-based face recognition it can also be edges or skin texture at different locations. Later holistic method gained popularity. Principle component analysis (PCA), independent component analysis (ICA) and linear discriminate analysis (LDA) work on the entire image to find reduced feature set. The whole image dimension is more, so computation required is more and is ineffective. Whereas if few features are extracted from the entire image then it helps in dimensionality reduction and computation wise it is more effective. PCA, ICA and LDA are the most popular traditional feature extraction algorithms for face representation.

The detail of principal component analysis (PCA) algorithm is as follows. A *mxn* image is a *mnx*1 feature vector and its dimension is *mn*. PCA gets rid of the curse of dimensionality in input images by removing redundant and highly correlated features. To reduce dimension PCA will choose few hyper-planes where if the data is projected, they will be maximally spread out and most of the information in the data will be retained.

First essential step for PCA is standardizing the pixel values by Z score standardization where $Z = \frac{feature\ value - mean}{standard\ deviation}$. Second step of PCA algorithm is computing the covariance matrix of the feature set to find their correlation. In the third step, Eigen values and Eigen vectors are computed from the covariance matrix to determine the principal component of the dataset. Eigen vectors are orthogonal or independent to each other and the number of Eigen vectors is same as the original image dimension. From the descending ordered set of Eigen vectors first few Eigen vectors are kept which retains maximum information of the original image and they are called the principal components. From the principal components a feature matrix is computed. Eigen vectors are also called Eigen faces as they resemble the original face. Independent component analysis is a generalization of PCA.

After this face representation with reduced feature sets, any supervised classification algorithms like *K*-nearest neighbour or multi-class support vector machine (SVM) can be used for face matching.

III. Modern Approach: Unified Embedding from Deep Learning

When the database size is large scale, performing face verification and recognition is a challenging task with traditional machine learning techniques. Also, different orientation of the face image is a challenging task for traditional methods. Modern deep learning techniques are more successful [14], [15], [16], [17], [18], [19] in handling adverse image quality and large dataset. In comparison to traditional methods of PCA, ICA and LDA which are linear mapping from input to output, artificial neural networks are nonlinear mapping techniques resulting in better accuracy for face recognition.

Deep learning methods are becoming popular with the increased computing power of GPUs and TPUs. Deep ANNs can give better accuracy when trained with huge amount of input data at the cost of computation time and speed. Since 2000 graphics processor unit (GPU) and since 2016 TensorFlow processor unit (TPU) are offering massive parallel processing hardware architecture for ANN based machine learning algorithms. Google has released TensorFlow computing platform that makes easy implementation of different artificial neural nets architecture with few lines of codes in programming languages like Python or R. So, like all other applications of computer vision and pattern recognition, automatic face recognition has transitioned to deep learning techniques.

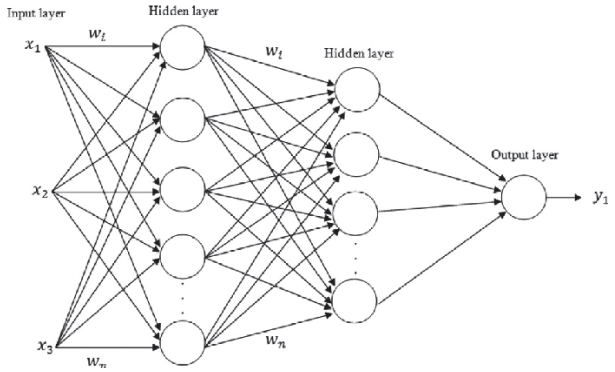

Fig. 4. Figure 4: MLP Classifier

Figure 4 shows a simple ANN architecture named multi-layer perceptron (MLP). ANNs are function approximator and can produce any function at the output to the desired accuracy. This makes a robust system to include real world input face image variations. In general, deep learning neural net is trained with a lot of data for several epochs to capture the best features that represent and generalizes the data. The purpose of the training is to create a model that generalizes the data enough for future prediction. As the number of epochs is increased the model accuracy increases up to some point for validation data set but then it starts declining due to overfitting of data during training. The number of epochs to train ANN thus can be determined by visualizing the model performance (accuracy vs. epoch no.) as shown in Figure 5. ANN architecture is optimized by adjusting its parameters such as number of nodes in each layer and adjusting number of layers that gives best performance. This procedure is also achieved by plotting model performance graph for training set data and validation set data.

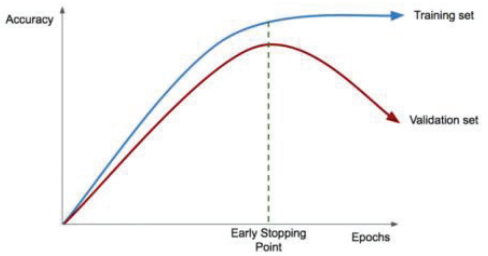

Fig. 5. Visualization of Model Performance

Authorized licensed use limited to: Deakin University. Downloaded on July 26,2021 at 05:43:21 UTC from IEEE Xplore. Restrictions apply.

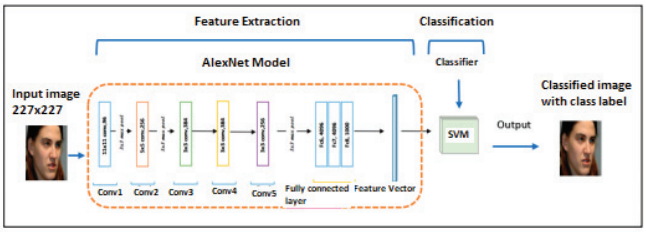

Fig. 6. AlexNet Convolution Neural Network with SVM Classifier for Face Recognition [20]

In deep learning, convolutional neural network (CNN) architecture [19], [20], [21], [22] as shown in Figure 6, has made remarkable progress in object detection, optical character recognition, face recognition and facial expression analysis. AlexNet designed in 2012 is a pre-trained CNN model for feature extraction or face representation with selected feature vector. This is also called face embedding in reduced feature space. AlexNet utilized fast GPU processing of the CNN and won the ILSVRC competition in 2012, in image recognition, both in terms of speed and accuracy. After faces are represented as feature vectors a supervised classification algorithm like *K*-nearest neighbour or multi-class support vector machine (SVM) algorithm is applied for face recognition.

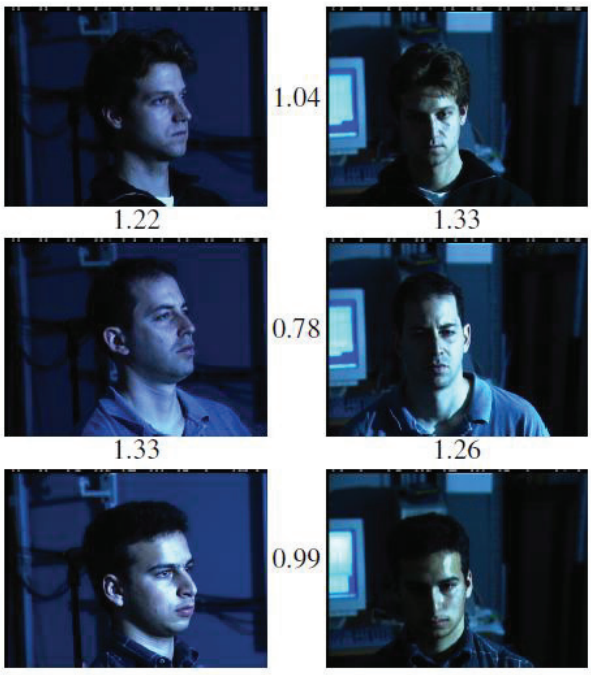

Fig. 7. CNN Output Distance Between Pair of Images of Same (different columns) and Different Person (different rows) [21]

In 2015 Google Inc. presented a deep convolution neural net (CNN) architecture named FaceNet [21] which performed 30% better in LFW and YouTube database than the state-of-the-art technique of that time. FaceNet maps each input thumbnail face image to a 128-D output embedding feature vector in Euclidian space requiring only 128 Bytes. This is a lot of dimension reduction for computing image to image similarity or dissimilarity measurements. The training of the model has the goal to have pose and illumination invariance in the output. At the input side, triplet of three images is used. Two images are slightly different alignment of the same person and a third image is of a different person. The goal is to produce output feature vectors, so that two images of same person will be close to each other than with the different person. Figure 7 shows FaceNet CNN output differences between images of same person (in two columns) and between different persons (in two rows).

FaceNet training method minimizes a loss function **L** of triplet dataset as given in Equation 1. Here L2 distance between two faces of the same person is minimized and distance between two different persons is at least some threshold value α. In Equation 1, the embedded output vector is represented as *f(x)* and $f(x_i^a)$, $f(x_i^p)$, *and* $f(x_i^n)$ are the anchor image, positive image of the same person and negative image of a different person in the triplet set.

$$\sum_{i=0}^{N}\left[\left\|f(x_i^a)-f(x_i^p)\right\|_2^2-\left\|f(x_i^a)-f(x_i^n)\right\|_2^2+\alpha\right]_+ \quad \text{.....} \quad (1)$$

Once the model is trained with large number of input dataset; for face verification to see whether two images are of the same person, Euclidian distance can be computed. If the distance is less than a threshold value 1.1, then two images are same otherwise not. Face recognition is done by finding K-nearest neighbour machine learning supervised classification algorithm. The model takes 2000 hours of training on CPU clusters to converge. FaceNet achieves very high accuracy rate of more than 99% with face alignment. Face recognition problem is solved for still images in LFW and video images in YouTube database with good quality and with clear Face alignment. But problem of poor-quality video surveillance images with all variable conditions is still challenging.

FaceNet was a deterministic face recognition model where no uncertainty or confidence measures is associated with face recognition. But in practice, if a human is asked to identify an image, especially a blurry image, he will attach an uncertainty level with it. For a completely corrupt image he will say it is not possible to identify the face because of its poor quality. So, in this paper [23] authors have come up with a model for probabilistic face embedding (PFE) technique. PFE gives a distributional estimate of features instead of a point estimate. A blurry image will be falsely rejected in deterministic embedding. In probabilistic encoding it will be recognized with some uncertainty measure. If the span of the distribution for a blurry image is very spread out it can even be not considered while training the model. During face recognition the model can penalize the uncertain feature vectors and can work with more confident feature vectors only. PFE uses maximum likelihood score for face matching. PFE performs better than deterministic methods in all face databases like LFW, YTF, IJB-A, IJB-C and IJB-S [7].

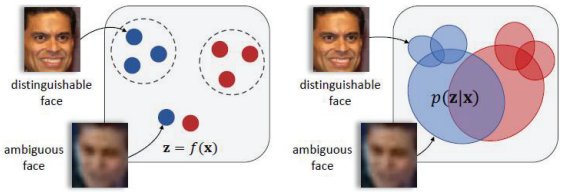

Fig. 8. a) Deterministic Embedding b) Probabilistic Embedding

Even though deep learning is better than traditional machine learning techniques, one thing to mention here is

Authorized licensed use limited to: Deakin University. Downloaded on July 26,2021 at 05:43:21 UTC from IEEE Xplore. Restrictions apply.

that individual researchers do not get access to behemoth of computing powers of GPUs and TPUs, like the technology giants Google, Amazon, Microsoft and FaceBook to research on deep learning techniques with large datasets. So, the traditional machine learning techniques for image processing is still going to exists.

IV. CONCLUSION

Face recognition is a preferred biometric identification method for security and authentication applications. The purpose of the paper is to present a comprehensive summary of improvements in automatic face recognition techniques in the last three decades. In the big data era, face recognition like all other applications of computer science has also transitioned into deep learning artificial neural network techniques to capture non-linear relation between input and output data for improved accuracy of the system. Achieving human level of accuracy in automatic face recognition in many real time applications with adversarial image quality will keep the researchers busy for many years to come. But only technology giants like Google, Amazon, Microsoft and FaceBook etc. can benefit from the parallel computing in GPUs and TPUs to research on deep learning techniques. So, the individual researchers will still depend on traditional machine learning techniques for image processing.

REFERENCES

[1] A. J. Mansfield et al. Best Practices in Testing and Reporting Performance of Biometric Devices. National Physics Laboratory Report CMSC 14/02, ISSN 1471-0005.

[2] M. A. Turk et al. Face Recognition Using Eigenfaces. 1991.

[3] Kresimir Delac et al. Recent Advances in Face Recognition. Published by In-Teh, www.in-teh.org, 2008.

[4] Phillips P. J. Support Vector Machines Applied to Face Recognition. Advanced Neural Information Processing System, 11, pp 803-809, 1998.

[5] K. Jonsson, J. Kittler, Y. Li, and J. Matas, Support Vector Machines for Face Authentication, Image and Vision Computing, Vol. 20, no. 5-6, pp. 369–375, 2002.

[6] Yandong Guo et al. MS-Celeb-1M: A Dataset and Benchmark for Large Scale Face Recognition. In Proceedings of European Conference of Computer Vision (ECCV) Workshops, 2016.

[7] Cameron Whitelam et al. IARPA Janus Benchmark-B Face Dataset.

[8] Kailash J. Karande et al. Laplacian of Gaussian Edge Detection for Face Recognition Using ICA. Springer, India, 2013.

[9] Paul Viola et al. Rapid Object Detection Using a Boosted Cascade of Simple Features. Proceedings on the IEEE Computer Society Conference of Computer Vision and Pattern Recognition. 2001.

[10] Navneet Dalal et al. Histograms of Oriented Gradients for Human Detection. IEEE Conference on Computer Vision and Pattern Recognition, 2005.

[11] J. Shi, A. Samal, and D. Marx, How Effective are Landmarks and Their Geometry for Face Recognition? Computer Vision and Image Understanding, Vol. 102, no. 2, pp. 117–133, 2006.

[12] L. Sirovich and M. Kirby, Low-Dimensional Procedure for the Characterization of Human Faces, Journal of Optical Society of America A, Vol 4, no. 3, pp. 519–524, 1987.

[13] Daniel Saez Trigueros et al. Face Recognition: From Traditional to Deep Learning Methods. arXiv:1811.00116v1 [cs.CV] 31 Oct 2018.

[14] K. Simonyan et al. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv Preprint arXiv:1409.1556, 2014.

[15] Sachin Sudhakar et al. Multi-View Face Detection Using Deep Convolutional Neural Networks. arXiv:1502.02766v3[cs.CV] Apr. 2015.

[16] Fei Wang et al. The Devil of Face Recognition is in the Noise. arXiv:1807.11649v1 [cs.CV] 31 Jul 2018

[17] Y. Sun, X. Wang, and X. Tang. Deeply Learned Face Representations are Sparse, Selective, and Robust. CoRR, abs/1412.1265, 2014. 1, 2, 5, 8

[18] Yaniv Taigman et al. DeepFace: Closing the Gap to Human-Level Performance in Face Verification

[19] Christian Szegedy et al. Going Deeper with Convolutions. Google Inc. arXiv:1409.4842v1 [cs.CV] 17 Sep 2014.

[20] Soad Almabdy et al. Deep Convolutional Neural Network-Based Approaches for Face Recognition. Applied Science, Vol. 9, 2019.

[21] Florian Schoff et al. FaceNet: A Unified Embedding for Face Recognition and Clustering. Google Inc. arXiv:1503.03832v3 [cs.CV] 17 Jun 2015.

[22] Y. Yamada, M. Iwamura, and K. Kise, Deep Pyramidal Residual Networks with Separated Stochastic Depth, arXiv Preprint arXiv:1612.01230, 2016.

[23] Yichun Shi et al. Probabilistic Face Embeddings. arXiv:1904.09658v4 [cs.CV] 7 Aug 2019.

Authorized licensed use limited to: Deakin University. Downloaded on July 26,2021 at 05:43:21 UTC from IEEE Xplore. Restrictions apply.