# Comprehensive Study on Real-time Video based Face Identification

**Patel Bhautika R.**

**Smt. Tanuben and Dr. Manubhai Trivedi College of Information Science, Surat, India.**

**bhautika.patel@gmail.com**

*Abstract* — **Nowadays, Security in the society is of high concern. To provide this we have seen a tremendous development in biometric based security system including face, iris, ear, odor, DNA, palm and many more. Each of these biometry is having its own pros and cons. Image based face recognition is still facing some challenges like lack of training images, illumination, pose variation, occlusion, distance etc. So to develop more robust face identification system we can collect spatial-temporal information from the video. Hence, in this paper a review of video-based face identification under different scenario is done to justify its need. The purpose of this is to get an idea of well-known methods used in this field. Finally the review showcase the power of Convolutional Neural Network (CNN), a branch of deep learning which is becoming more popular for analyzing images in the field of computer vision.**

***Keywords*— *AdaBoost, Principal Component Analysis (PCA), Computer Vision,* Convolutional Neural Network (CNN), Face Recognition.**

## I. INTRODUCTION

The usefulness of every biometric system including iris, DNA, face or any other depends on the circumstances in which it is used. Face is one of the most non-intrusive biometry of a human being that tells a lot about oneself and used widely for identification nowadays. Face identification can be done from the still image as well as from video using the techniques of computer vision. But identification from video is becoming more popular as it gives abandon information about the person with time to time and hence helps the researchers to improve accuracy and robustness. Although it has many challenges like large scale variations, low quality of facial images, illumination changes, pose variations, occlusions etc. Apart from this face changes with factors like moustache, beard, glasses, size as well as facial expression.

The research on the video-based face identification is an extension of the still-image face identification, which has been widely researched for years and some excellent results have been reported in the literature. According to [1] there are three distinct scenarios for video-based face recognition system, (i) Video-to-Still (V2S); (ii) Still-to-Video (S2V); and (iii) Video-to-Video (V2V), respectively, taking video or still image as query or target. Video-based face identification can be done by capturing the video segments, performing preprocessing, detecting the face, extracting face features and then classifying the face.

There are many methodologies available for face identification but the literature shows that machine learning is more promising. Steps that can be applied while performing identification using machine learning is present in figure 1.

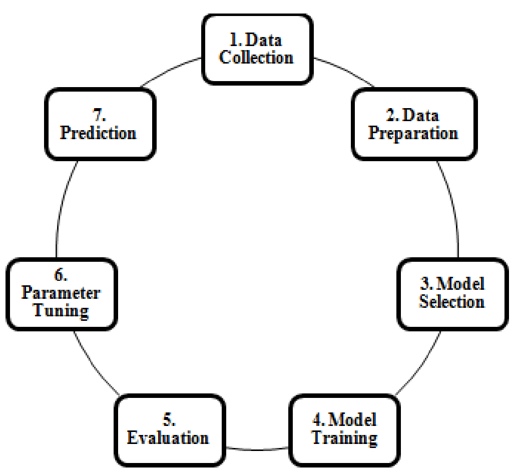

**Figure 1. Steps used in Machine Learning Application**

During data collection stage, we can collect the facial data from still images, available videos' as well as from the real-time video. For the machine learning applications we need humongous amount of data for training, evaluation and prediction. This step is very critical. The quality and the quantity of data collected at this stage have direct impact on model prediction. Some large scale face datasets are LFW Face dataset, MS-Celeb dataset, Mega Face dataset, VGG Face dataset.

Next step is data preparation. At this stage activities like data normalization, de-duplication, error correction; labelling etc is done. We should divide filtered data into

DOI : 10.35291/2454-9150.2020.0590

three datasets. (i) Training dataset (ii) Evaluation dataset and (iii) Testing dataset. Each of the dataset must not be repeated and should be selected randomly. We can also so some sort of data visualization to see the relationship between variables we have collected.

After this, we have to select the appropriate machine learning model based on our problem. Some of the well-known machine learning models is ANN, CNN, RNN etc.

After the model selection, now train the model using training data set. The purpose of this step is to gradually improve model ability to improve face recognition. Never ever use the testing data to train the model. As far as face identification model is concern the model is trained by applying steps shown in figure 2.

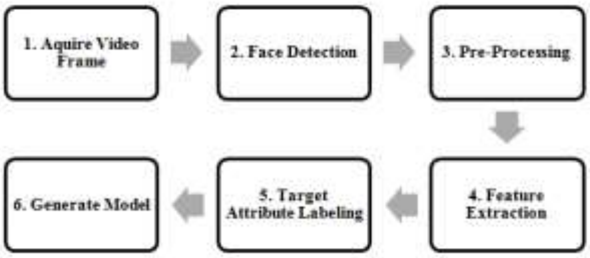

**Figure 2. Model Training Stages**

After model building next evaluate it on the evaluation dataset using some evaluation matrices like confusion matrix.

Hyper-parameter tunning is required in order to remove over-fitting and under-fitting errors and allow us to select the right parameters.

Finally we can deploy our model and perform the prediction using the test dataset.

## II. LITERATURE REVIEW

Video based face recognition has attracted a great deal of attention in last few years. Many researchers have developed a promising algorithm including face recognition using multiple classifiers such as majority voting and sum rule [2] where the algorithm was tested on the XM2VTS face video database and they achieve good recognition accuracy. Moreover the database was captured under high quality digital video.

A video-based face detection and recognition system was developed by Ping Zhang [3]. Methods used for face detection includes face skin, face symmetry and eye symmetry verification and the detection rate is ~95%. Three layers ANN with Back-Propagation (BP) was used to classify the image.

Changbo Hu et. al [4] proposed patch-based face recognition from video and the model is tested against still images as well as on video sequence and the recognition rate was 81%.

Face detection and recognition from video sequence that uses Kalman filter for face tracking, AdaBoost algorithm for face detection and Pseudo-two-dimensional Hidden Markov Model for feature extraction and recognition [5]. The algorithm was tested on six subjects and the frame rate was 25 frames per second.

Locality Preserving Projections (LPP) based face recognition method from video was developed by [6] that used k-nearest neighbor. The algorithm was tested on 18 different persons of Honda/UCSD Video Database, including 18 training videos and 18 testing videos.

Le An et al [7] developed face recognition system from multi-camera surveillance videos that fused the faces from two cameras to improve the recognition result. They have proposed video based face recognition model using a Unified Face Image (UFI) that is produced from multiple camera. The classification is done using K-Nearest Neighbor (KNN) with Chi-Square distance. The model was tested on ChokePoint dataset that contains only frontal faces of the 25 individual that too is with high resolution and the highest recognition rate was 48%.

Video-based face recognition that uses image averaging was proposed by Peng Jia and Dewen Hu [8] to improve face recognition accuracy. The algorithm was tested on Honda/UCSD video database consists of 38 video sequences of 19 different subjects, 2 videos per subject, one for training and the other for testing. They have used Principal Component Analysis (PCA) and Locally Linear Embedding (LLE) of dimensionality reduction and Nearest Neighbor for classification.

A semantic model for video based face recognition was proposed by Dihong Gong et. al [9] and have extracted key frames from each video sequence using the spatio-temporal synchronization method. PCA+LDA were used for feature representation and have used k-means for semantic classification. The model was tested on XM2VTS video database and the identification accuracy is 95.33% whereas verification rate is 97.82%.

Zhiwu Huang et. al [1] developed video-based face recognition system that was evaluated on benchmark face dataset COX. The algorithm was evaluated for all the three video-based face recognition scenarios: V2S, S2V and V2V and the best identification rates for V2S/S2V scenarios was around 40%~60%, while those for V2V scenario was around 75%~ 85%.

Face recognition from video sequence using fuzzy approach and multi-agent was developed by Hicham Hatimi et al [10]. They have used fuzzy approach for face detection and recognition using multi-agent modeling to find the degree of membership. The system was evaluated on ORL (Olivetti Research Laboratory) face image database containing 400 images of 40 individuals. The results for

DOI : 10.35291/2454-9150.2020.0590

two tests were 95% and 92% respectively. But the overall algorithm accuracy was not present.

Human faces in surveillance videos often suffer from image blur, dramatic pose variations, and occlusion. Hence Video-based face recognition using Convolution Neural Network (CNN) as well as Trunk-Branch Ensemble CNN model (TBE-CNN) was proposed by [11] to overcome these challenges. They have artificially blurred training data composed of clear still images to account for a shortfall in real-world training data. The performance of the algorithm was evaluated on three popular video face databases: PaSC, COX Face, and YouTube Faces and achieved good performance.

In [12], autoregressive and moving average (ARMA) model was used for face recognition that take care of the changes in appearance and recognition was performed using subspace angle. The model was tested on two different datasets; the first dataset has face videos for 16 subjects with 2 sequences per subject and the dataset was UCSD/Honda. They got recognition performance of more than 90% (15/16 for the first dataset and 27/30 for the second).

Face recognition in real-world surveillance using deep learning model was proposed by Ya Wang et. Al [13]. They have used eight convolutional layer and three fully connected layer followed by one or more non-linearities such as ReLU and max-pooling. The model was tested after fine-tuning the VGG face dataset and the recognition accuracy for 140 and 240 identification classes was 91.4% and 92.1% respectively.

Face recognition from video using adaptive Hidden Markov Model (HMM) [14] has been proposed by some researchers. The model was tested against Mobo and Task dataset with the videos of 24 and 21 subjects and they obtain 1.2% and 4.0% recognition error rate respectively.

Face recognition from video using real world data has been proposed by Johannes Stallkamp et. al [15], which recognized the people entering through some laboratory. They have used three different measures to weight the contribution of each individual frame for classification; those are distance to-model (DTM), distance-to-second-closest (DT2ND), and their combination. Individual frame score is calculated using Gaussian mixtures and finally classification is done using KNN with value of k=10. The algorithm was tested on database of 41 subjects and after combined approach the recognition result was 91.8%.

Face recognition from video sequence using probabilistic appearance manifolds has been developed by Kuang-Chih Lee [16] et al. they have introduces weight mask in order to recognize face that are partially occluded. The algorithm was tested on set of 45 videos of 20 different people; each individual in our database has at least two videos where each person moves in a different combination of 2-D and 3-D rotation, expression, and speed. Each video was recorded in an indoor environment and each one lasted for at least 20 seconds. The recognition result without occlusion was 95.1% and with occlusion it was 93.3%. Though the proposed model handles large motions well, it is nevertheless sensitive to large illumination changes.

Face recognition using probabilistic appearance manifolds has also been proposed by Kuang-Chih Lee [17] et al. in order to train and test the model they have collected a set of 52 video sequences of 20 different persons for the task of testing our system. Each video sequence is recorded by a SONY EVI D30 camera in an indoor environment at 15 frames per second, and each lasted for at least 20 s. The resolution of each video sequence is 640 X 480.out of 52 video sequences 20 were used for training whereas rest was used for testing. They have manually cropped the images from the video frames which are further down-sampled to a standard size of 19 X 19 pixels. The video sequence also includes some partially occluded faces and achieved better performance. The dataset is available for download at url [18].

An automatic system for unconstrained video-based face recognition has been developed by Jingxiao Zheng [19] et al. They have used Deep Convolutional Neural Networks (DCNN) , which is has impressive object recognition capability and used in computer vision. The system was trained and tested in four databases namely IARPA Benchmark B (IJB-B), ARPA Janus Surveillance Video Benchmark (IJB-S), Multiple Biometric Grand Challenge (MBGC) dataset and the Face and Ocular Challenge Series (FOCS) dataset. Among these IJB-B and IJB-S datasets are captured in unconstrained settings and contain faces with much more intra/inter class variations on pose, illumination, occlusion, video quality, scale and etc. They have achieved good performance on MBGC and FOCS datasets compared to other using different methods.

Recognition of face from face motion manifolds that uses robust Kernel Resistor-Average Distance (K-RAD) has been proposed by Ognjen Arandjelovic and Roberto Cipolla [20]. The algorithm was tested on 6 different datasets of around 35-90 people and the average accuracy was 98%. Although clear idea of these datasets is not present, they have captured 30-50 images of each person in random motion. Further lighting condition for each dataset varies and images we reduces to 15x15 pixel gray scale image to reduce computational cost.

Ognjen Arandjelovic and Roberto Cipolla [21] have developed video base face recognition model named pose-wise linear illumination manifold. The system was capable to recognize faces in unconstrained environment under variety of pose, lighting and even with low image resolution. To improve recognition under varied illumination they have used coarse histogram correction along with illumination manifold-based normalization. In

DOI : 10.35291/2454-9150.2020.0590

order to recognize face against pose variation they have decomposed each image into semantic Gaussian pose clusters. The model was evaluated on FaceDB60 dataset, having faces of 60 individuals of different age. Further most faces were Japanese and male. The average accuracy of this model was 95%.

Face recognition using multiple pose-aware deep learning models has been introduced by Wael Abd Almageed et al [22]. In this system, a face image is processed by several pose-specific deep convolutional neural network (CNN) models to generate multiple pose-specific features which reduce pose variation. 3D rendering is used to generate multiple face poses from the input image.

Face recognition system that uses Convolution Neural Network (CNN) to detect the facial images and training, and Logistic Regression Classifier (LRC) to classify the features learned by CNN has been proposed by Hurieh Khalajzadeh et al [23]. They have applied CNN composed of four layers: input layer, two convolutional layers and one sub-sampling layer. The output layer is a fully connected layer with 15 feature maps. The Yale face database contains 165 face images of 15 individuals has been used to evaluate the performance of the system, there are 11 images of each individual, and out of that 9 were used for training and 2 for testing. The accuracy of this system was 83.63%. Although, the size of dataset was not sufficient.

Classification on ImageNet database using deep convolutional neural network has been developed by Alex Krizhevsky [24] et al. The CNN adopted by them have five convolutional layers some of which are followed by max-pooling layers and three fully connected layers. To reduce over-fitting in the fully-connected layers they employed a recently-developed regularization method called “dropout” that proved to be very effective. The ratio of training and testing images from ImageNet dataset was 50-50. The result on this database was 67.4%.

Convolutional Neural Networks (CNNs) have been established as a powerful class of models for image recognition problems. So large-scale video classification with CNN has been approached [25]. They have used Sports-1M dataset, which consists of 1 million YouTube videos belonging to taxonomy of 487 classes of sports to train and test the model. They have used Multi-resolution CNN to speed up the performance. They split the dataset by assigning 70% of the videos to the training set, 10% to a validation set and 20% to a test set. The average accuracy of this model was 82.4%.

Deep learning has led to a very good performance on different problem domain like speech recognition, visual recognition and natural language processing. Jiuxiang Gu [26] et al introduced recent advances in convolutional neural networks with various application domains of CNN. According to them LeNet-5 was the first CNN followed by other evolutions of CNN like AlexNet, ZFNet, VGGNet, GoogleNet and RasNet. There have been various improvements on CNNs since the success of AlexNet in 2012. These improvements are from six aspects: convolutional layer, pooling layer, activation function, loss function, regularization, and optimization. Further they also highlighted some problem areas where CNN is recently used to achieve state-of-art performance including image classification, object tracking, pose estimation, text detection, visual saliency detection, action recognition, scene labeling, speech and natural language processing.

With the research and development of the deep learning method, convolution neural network and many other excellent machine learning methods have emerged, which have made breakthrough progress in many applications, such as image recognition, target classification and so on. Guifang Lin and Wei Shen [27] have proposed a model based on CNN that uses CIFAR-10 data set using TensorFlow library for image classification. They have adopted different activation functions like SignReLu, ReLu, ReLu6 and Elu and concluded that maximum recognition rate is 86.96% using SignReLu activation function.

Convolutional neural networks (CNN) have proven to be very successful to analyzing visual imagery. Different types of CNN architectures have been developed to solve different problems. Some well known CNN for face recognition are: CASIA WebFace, FaceNet and VGGFace2 which are computationally very expensive. So, the binary version of CNN named Local Binary CNN (LBCNN) has been proposed by Carolina Toledo Ferraz and José Hiroki Saito [28]. The advantage of this architecture is that it has lower model complexity and is less prone to over-fitting. They have also presented algorithm accuracy using noisy image only in test as well as training and testing and found better result in second scenario.

## III. CONCLUSION AND FUTURE WORK

In this paper, a review of well-known video-based face identification algorithm is conducted. Each uses different approach for feature extraction and classification. Some of the algorithms used for classification that are reviewed in this paper includes ANN, LPP, PCA, LLE, PCA+LDA, Fuzzy logic, CNN with different variations, HMM, DCNN etc.

The study shows that amongst the various available methods Convolutional Neural Network (CNN) is dominating. It is widely used as well as the recognition accuracy using this approach is better than others at the same time it requires large amount of data to train the model. It also highlights the basic steps used in machine learning as well as various layers used in CNN like Input layer, hidden layer and output layer. Basic steps required in order to perform video-based face recognition is present so that a novice user can effectively understand. Even

DOI : 10.35291/2454-9150.2020.0590

beginners can get to know about available facial datasets through this paper. Once identification is done then in future we can track the person by analyzing video frames.

## REFERENCES

[1] Zhiwu Huang, Shiguang Shan, Ruiping Wang, Haihong Zhang, Shihong Lao and Xilin Chen (2015). A Benchmark and Comparative Study of Video-Based Face Recognition on COX Face Database. IEEE transactions on image processing, Volume 24.

[2] Xiaoou Tang and Zhifeng Li (2004). Video Based Face Recognition Using Multiple Classifiers. Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR'04).

[3] Ping Zhang (2008), A Video-based Face Detection and Recognition System using Cascade Face Verification Modules, 37th IEEE Applied Imagery Pattern Recognition Workshop.

[4] Changbo Hu, Josh Harguess and J. K. Aggarwal (2009). Patch-Based Face Recognition from Video. 16th IEEE International Conference on Image Processing (ICIP).

[5] Qianqian Zhao and HualongCai (2010). The Research and Implementation of Face Detection and Recognition Based on Video Sequences, 2nd International Conference on Future Computer and Communication.

[6] Ke Lu, Zhengming Ding, Jidong Zhao and Yue Wu (2010). Video-based Face Recognition. 3rd International Congress on Image and Signal Processing (CISP2010).

[7] Le An, BirBhanu and Songfan Yang (2012). Face Recognition in Multi-Camera Surveillance Videos. 21st International Conference on Pattern Recognition (ICPR 2012).

[8] Peng Jia and Dewen Hu, Video-Based Face Recognition Using Image Averaging Technique, 5th International Congress on Image and Signal Processing (CISP 2012).

[9] Dihong Gong, Kai Zhu, Zhifeng Li, and Yu Qiao (2013). A Semantic Model for Video Based Face Recognition, Proceeding of the IEEE International Conference on Information and Automation, China.

[10] Hicham Hatimi, Mohamed Fakir and Mohamed Chabi (2016). Face recognition using a fuzzy approach and a multi-agent system from video sequences. 13th International Conference Computer Graphics, Imaging and Visualization.

[11] Changxing Ding and Dacheng Tao (2018). Trunk-Branch Ensemble Convolutional Neural Networks for Video-Based Face Recognition, IEEE transactions on pattern analysis and machine intelligence, Volume 40.

[12] Gaurav Aggarwal, Amit K. Roy Chowdhury and Rama Chellappa (2004). A System Identification Approach for Video-based Face Recognition. Proceedings of the 17th International Conference on Pattern Recognition.

[13] Ya Wang, Tianlong Bao, Chunhui Ding and Ming Zhu (2017). Face Recognition in Real-world Surveillance Videos with Deep Learning Method. 2nd International Conference on Image, Vision and Computing.

[14] Xiaoming Liu and Tsuhan Chen (2003). Video-Based Face Recognition Using Adaptive Hidden Markov Models. Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03).

[15] Johannes Stallkamp, Hazım K. Ekenel and Rainer Stiefelhagen (2007). Video-based Face Recognition on Real-World Data, IEEE-2007.

[16] Kuang-Chih Lee, Jeffrey Ho, Ming-Hsuan Yang and David Kriegman (2003). Video-Based Face Recognition Using Probabilistic Appearance Manifolds. IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[17] Kuang-Chih Lee, Jeffrey Ho, Ming-Hsuan Yang and David Kriegman (2005). Visual tracking and recognition using probabilistic appearance manifolds. Computer Vision and Image Understanding 99 (2005) 303–331.

[18] http://vision.ucsd.edu/kriegman-grp/research/vfr/.

[19] Jingxiao Zheng, Rajeev Ranjan, Ching-Hui Chen, Jun-Cheng Chen, Carlos D. Castillo and Rama Chellappa (2020). An Automatic System for Unconstrained Video-Based Face Recognition. IEEE Transactions on Biometrics, Behavior, and Identity Science ( Volume: 2 , Issue: 3).

[20] Ognjen Arandjelovi´c and Roberto Cipolla (2004). Face Recognition from Face Motion Manifolds using Robust Kernel Resistor-Average Distance. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'04).

[21] Ognjen Arandjelovi´c and Roberto Cipolla (2008). A pose-wise linear illumination manifold model for face recognition using video. Computer Vision and Image Understanding.

[22] Wael AbdAlmageed and Yue Wu et al (2016). Face recognition using deep multi-pose representations. IEEE Winter Conference on Applications of Computer Vision (WACV).

[23] Hurieh Khalajzadeh, Mohammad Mansouri and Mohammad Teshnehlab (2012). Face Recognition using Convolutional Neural Network and Simple Logistic Classifier. Online conference on Soft Computing in Industrial Applications.

[24] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton (2012). ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems 25 (NIPS 2012).

[25] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar and Li Fei-Fei (2014). Large-scale Video Classification with Convolutional Neural Networks. IEEE Conference on Computer Vision and Pattern Recognition.

[26] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Li Wang, Gang Wang, Jianfei Cai and Tsuhan Chen (2017). Recent Advances in Convolutional Neural Networks. ArXiv:1512.07108v6.

[27] Guifang Lin and Wei Shen (2018). Research on convolutional neural network based on improved Relu piecewise activation function. 8th International Congress of Information and Communication Technology (ICICT-2018).

[28] Carolina Toledo Ferraz and José Hiroki Saito (2018). A comprehensive analysis of Local Binary Convolutional Neural Network for fast face recognition in surveillance video. In Brazilian Symposium on Multimedia and the Web (WebMedia '18), Salvador-BA, Brazil. ACM, New York, NY, USA.

DOI : 10.35291/2454-9150.2020.0590