Face Recognition from Video: An overview

PROF.KALAM NARREN
Department of Computer Engineering
Rajarshi Shahu College of Engineering
Pune, India

PROF.V.VINAY KRISHNA
Department of Computer Engineering
Rajarshi Shahu College of Engineering
Pune, India

ABSTRACT

Face recognition system involves detection and recognition of face from digital media i.e. image or video and has multi-domain applications from social media to security. Wide range of techniques have been applied over last few years to perform face recognition from still image. However, face recognition from video is still good area of research in terms of accuracy and performance. This survey paper highlights some common challenges faced in real time face recognition from video. Some well-known face recognition systems are discussed which uses different combination of techniques for face detection and face recognition. This paper also reviews pros of these face recognition systems and challenges it overcomes.
Keywords—Face Recognition from Video, Challenges, Methodologies.

I. INTRODUCTION

Face recognition in simple terms is identifying a person from digital media i.e. still image or video sequence. Face recognition from images has been widely studied in past few years to achieve impressive performance. Recently, the ubiquitous use of video capturing devices is shifting the focus of face recognition research from image-based scenarios to video-based ones [2]. Widespread use of video cameras for surveillance and mobile devices has led to enormous quantity of video being captured constantly. Compared to still face images, videos usually contain more information like temporal and multi-view information [3]. All these facts makes face recognition from video an interesting area of research.

There are variety of applications which extends basic face recognition technique in multiple disciplines. In security domain, it is highly desirable to build surveillance systems coupled with face recognition techniques to automatically identify subjects of interest. Security applications for electronic transactions and access control uses face recognition as biometric technique. Smart homes, human-machine interaction, interactive movies, computer games, image/ video search are few more systems using face recognition.

Human face is a flexible and deformable object, so the task of correctly modeling the prototype identity of an individual is very complex [9]. Face recognition from video is more challenging compared to that from still image because motion of the subject and the typical low quality of the image sensors makes the captured faces suffer from low resolution, noise, pose variation and blurriness together with unconstrained lighting conditions. Different state of the art methodologies have been proposed [1]-[14] to deal with these challenges.

In this paper, we provide critical review of challenges and current methodologies in face recognition. Section II introduces general model of face recognition and common challenges faced during face recognition from video. Section III gives an overview of few state of the art methodologies. Section IV has discussion and comparison of techniques discussed in earlier sections. Section V contains conclusive remarks.

II. GENERAL MODEL AND CHALLENGES

A. General Model of Face Recognition System

General model of face recognition system (FRS) is depicted in figure 1. It has following components-

- Input: FRS accepts image or video sequence (sequence of frames) as input from which face is to be recognized.
- Face Detection: Finds all faces in an image (if any) regardless of their position, scale, rotation, pose, illumination, facial expressions and occlusions. Exact location of detected face within image is calculated [20] [21]. When detected face is followed in video, it is called face tracking. Techniques for face detection can be categorised as[21]-
    - Knowledge based - Translate knowledge about typical face to a set of rules.

  - Template matching- Standard patterns are used to describing the face or facial features.
  - Feature invariant- Find features of the face invariant to appearance variations (facial features, edges, shape, texture, skin color)
  - Appearance based- Learn face characteristics from a representative set of example images using classic machine learning techniques
- Pre-processing: Noise removal and image normalization is done. This also includes face frontalization and face alignment.
- Feature Extraction: This is a form of dimensionality reduction wherein object of interest is represented as finite number of variables such that subsequent process of classification discriminates object correctly. Properly optimized feature extraction is the key to effective model construction.
- Classification: The goal of classification is to accurately predict the target class for each observation. Few examples of classifier are k-Nearest Neighbor, Hidden Markov Model, Naive Bayes Classifier, Support Vector Machines and Neural Network.

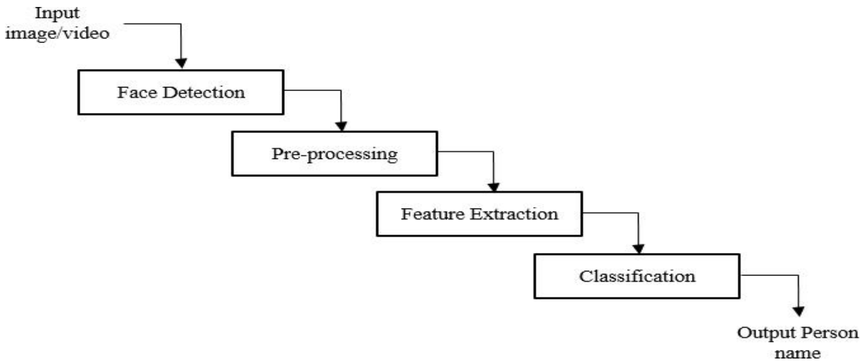

Figure 1: General model of Face Recognition System

Different FRS differ in method used for face detection, feature extraction and classification process.

B. **Challenges**

The real challenge in face detection and recognition technologies is the ability to handle all those scenarios where subjects are non-cooperative and the acquisition phase is unconstrained [22]. The performance of video FRS can decline significantly because reference still images are typically captured under controlled conditions in the enrolment domain (ED) with a still camera, and video frames are captured under uncontrolled conditions in operational domain (OD) with variations in pose, illumination, blurriness, etc. [5]. Two main categories of challenges are interclass similarity and intrapersonal variations.

In interclass similarity two different person look alike e.g. twins, relatives, strangers may also look alike. In intrapersonal variations same person may look differently due to illumination variations, pose variations, facial expressions, use of cosmetics and accessories, hairstyle changes, temporal variations (aging) etc.

A benchmark dataset for face detection named WIDER Face dataset with high degree of variability in scale, pose, occlusion, expression, appearance and illumination is proposed in [15] and shown in figure 2.

Figure 2: Challenges in Face Detection [15]

Other major challenges in case of surveillance videos are of low resolution and partial face recognition.

III. METHODOLOGIES

Multiple techniques are available for face detection, feature extraction and classification process. Combination of these techniques defines different FRS models. Few models are listed below:

- Patch based method: - A method to recognize a face from video based on face patches is proposed in [10]. First, face patches are cropped from the video frame by frame. Then, face patches are matched to an overall face model and stitched together. By accumulating the patches, a reconstructed face is built which is used in recognition. Recognition is done via sparse representation. Occlusion is handled by reconstructing face from patches.
- AdaBoost: - In [6] paper, author used the method of difference in background images and the Kalman filter to track and extract the region of human body firstly, and then used the AdaBoost algorithm to detect human face in the region. Finally the improved Hidden Markov Model which is named as Pseudo two-dimensional Hidden Markov Model is used for feature extraction and recognition in face image.
- Set based method: - In [2] author proposed a novel Point-to-Set Correlation Learning (PSCL) method for Video to still, still to video or video to video face recognition. PSCL method considers heterogeneity between still image (Euclidean point) and video sequence (Riemannian elements) and outperforms then state of the art methods.
- Laplacianfaces: - In [8] author used Laplacianfaces i.e. Locality Preserving Projections (LPP) to recognize video-based face sequence, so it can discover more space-time semantic information hidden in video face sequence, simultaneously make full use of the intrinsic nonlinear structure information to extract discriminative manifold features.
- Unified Face Image (UFI) Generation: - The poses of probe data are usually not frontal view, contrary to the standard format of the gallery data. Author of paper [7] proposed a multi-camera video based face recognition framework using a novel image representation called Unified Face Image (UFI), which is synthesized from multiple camera feeds. The probe frames from different cameras are wrapped towards a template frontal face and then averaged. The generated UFI is a frontal view of the subject. UFI generation from two cameras C1 and C2 is shown in figure 3. SIFT flow is used as a high level alignment tool to warp the faces. Recognition done using nearest neighbour classifier.

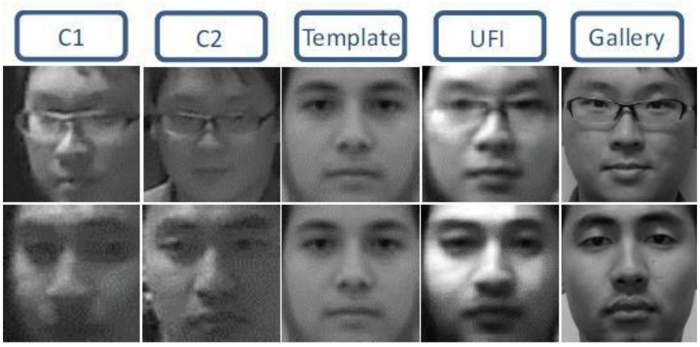

Figure 3: UFI Generation [7] [24]

- Cross Resolution FRS with Fully convolutional network (FCN): - Paper [1] is about cross resolution face identification where the enrolled face images are mostly collected in controlled scenarios with high resolution (HR) and low resolution (LR) faces are captured with surveillance cameras with an uncontrolled pose and lighting conditions. This is a challenging recognition task which relies strongly on a good resolution invariant representation. Here, it is expected to learn a common feature space that is able to cluster LR and HR faces within the same subject as well as maintaining low inter-class proximity despite the difference in resolution. Author proposed an approach which employs a novel regularization term to force the features from the same subject to cluster, which yields better discriminative features. Fully convolutional network (FCN) with spatial pyramid pooling (SPP) layer is used.
- Fuzzy method: - In [9], faces are detected using texture color and geometrical face. In the second step, the multi-agent system and fuzzy approach are used in the recognition process. The results obtained using this method demonstrates performance in terms of robustness, in the illumination variations and speed.
- K-means: - In [12] training faces from the same person are clustered semantically with the K-means algorithm, then faces from different persons are further grouped according to semantics. Finally, face features from each group are sliced into n slices, and PCA (Principal Component Analysis) + Linear

Discriminant Analysis (LDA) algorithm is applied yielding n*3 classifiers. At the testing stage, the matching score of the given pair of testing video sequences is computed by semantically assigning each frame into one of the groups. Final matching score is the sum of the best matching scores from each group

- Convolutional Neural Network: In [3], Trunk Branch Ensemble Convolution Neural Network TBE-CNN with improved triplet loss function is used to tackle occlusion, pose variation and blur images for Video based FR. TBE-CNN incorporates one trunk network and several branch networks. The trunk network is trained to learn face representations for holistic face images, and each branch network is trained to learn face representations for image patches cropped from one facial component. Figure 4 details architecture of TBE-CNN.

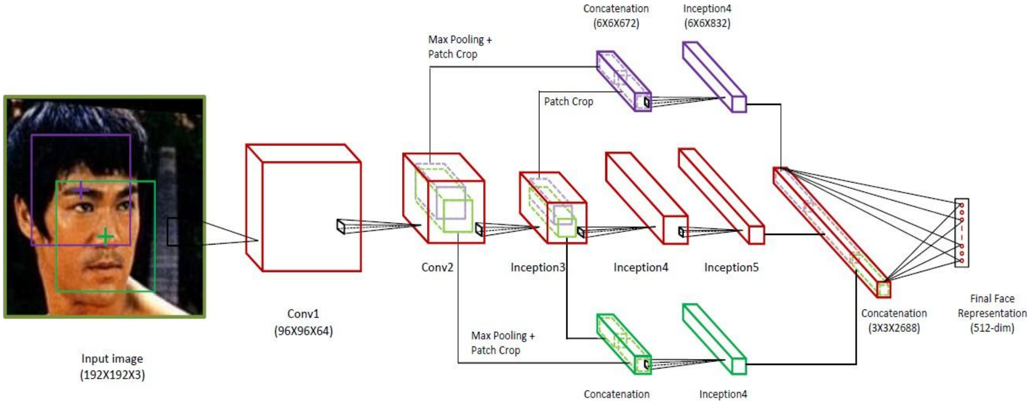

Figure 4: Model architechture of TBE-CNN [3]

- Transfer Learning: In [11], for face recognition author uses transfer learning to fine tune deep learning VGG face model [23] on surveillance dataset created automatically with little or no manual work.
- Sparse Representation: In [14], author proposed RVPose algorithm to handle pose variation. The key idea is performing alignment and recognition based on sparse representation simultaneously. RVPose considers that multi-pose faces possess the same texture and 3-dimensional shape. RVPose aligns a sequence of faces with pose variations simultaneously, which is reduced to a 3-dimensional shape-constrained video alignment problem.
- Dynamin Feature Matching: In [13], partial face recognition framework is proposed using dynamic feature matching (DFM), which could handle partial faces of arbitrary sizes without extra pre-processing with high accuracy and computational efficiency. Here, FCN is applied to extract spatial feature maps of given gallery and probe faces. Then sliding window of the same size as the probe feature maps is set to decompose the gallery feature maps into several gallery sub-feature maps. Finally, Sparse Representation Classification (SRC) is used to achieve alignment-free matching.
- Dataset creation: The performance of still-to-video FR systems can decline significantly when individual has very few or single reference image and that too with different capture conditions [5]. Domain specific face synthesis (DSFS) algorithm exploits intra class variation information to augment reference set by characterizing capture conditions. DSFS algorithm is explained in figure 5. Augmenting the reference gallery set of still-to-video FR systems using the proposed DSFS approach can provide a significantly higher level of accuracy compared to state-of-the-art approaches, with only a moderate increase in its computational complexity.

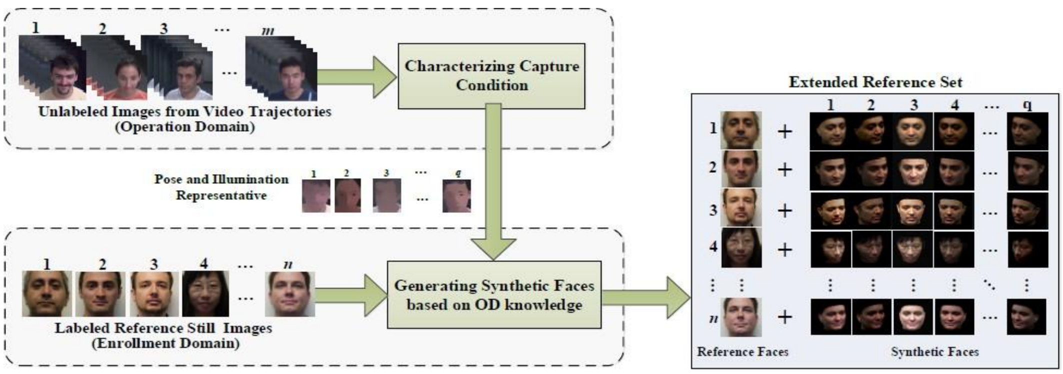

Figure 5: DSFS Algorithm [5]

IV. COMPARISON

Details of all methods listed above are compared in Table 1 for challenges it overcomes.

Table 1: Comparison of various approaches for video based FRS

Year	Source	Paper Title	Technique	Pros
2019	[1]	On Low-Resolution Face Recognition in the Wild: Comparisons and New Techniques	DAGAN	Handles low resolution
2019	[13]	Dynamic Feature Matching for Partial Face Recognition	FCN+ SRC	Improves performance of partial face recognition
2019	[5]	Domain-specific face synthesis for video face recognition from a single sample per person	DSFS + SRS	FR based on a single facial image
2018	[3]	Trunk Branch Ensemble Convolutional Neural Networks for Video-based Face Recognition	TBE-CNN with improved triplet loss	Deals with occlusion, pose, low resolution
2018	[14]	Robust Video Face Recognition Under Pose Variation	RVPose	Handles pose variation
2017	[11]	Face Recognition in Real-world Surveillance Videos with Deep Learning Method	Transfer Learning with VGG face model	Improves performance
2017	[4]	An Advancement towards Efficient Face Recognition Using Live Video Feed: For the Future	LDA+SVD	Higher efficiency and accuracy
2016	[9]	Face recognition using a fuzzy approach and a multi-agent system from video sequences	Multi Agent + Fuzzy	Handles variations illumination and speed
2015	[2]	A benchmark and comparative study of video-based face recognition on COX face database	PSCL	Outperforms most of the existing methods for V2S/S2V face recognition tasks
2013	[12]	A Semantic Model for Video Based Face Recognition	K- Means + PCA+ LDA	Great improvement over the traditional face recognition approaches
2012	[7]	Face recognition in multi-camera surveillance videos	SIFT LBP + K-Means	Handles pose variations and multi views
2010	[6]	The research and implementation of face detection and recognition based on video sequences	AdaBoost + Hidden Markov Model	Effective extract of motive human body in the video and then detection and recognition of face
2010	[8]	Video-based Face Recognition	LPP	Acquire more semantic information inside the video face
2009	[10]	Patch-based Face Recognition from video	Patch based method + Sparse representation	Handles noise and occlusion

V. CONCLUSION

Face recognition is very challenging process. Detailed studies show that video based face recognition is far from mature especially compared with face recognition from still images. More efforts should be made to handle general challenges in FRS along with achieving accuracy and performance.

ACKNOWLEDGEMENT

The authors would like to thank the researchers as well as publishers for making their resources available and teachers for their guidance. We are thankful to concern members of ICATCSIT-2019 organized by Rajarshi Shahu College of Engineering, Tathawade, Pune, for their constant guidelines and support. We are also thankful to the reviewer for their valuable suggestions.

REFERENCES

[1] Pei Li, Loreto Prieto, Domingo Mery and Patrick J. Flynn,"On low-resolution face recognition in the wild: comparisons and new techniques", DOI 10.1109/TIFS.2018.2890812, IEEE Transactions on Information Forensics and Security.

[2] Huang, Zhiwu, Shiguang Shan, Ruiping Wang, Haihong Zhang, Shihong Lao, Alifu Kuerban, and Xilin Chen. "A benchmark and comparative study of video-based face recognition on COX face database." IEEE Transactions on Image Processing 24, no. 12 (2015): 5967-5981.

[3] Ding, Changxing, and Dacheng Tao. "Trunk-branch ensemble convolutional neural networks for video-based face recognition." IEEE transactions on pattern analysis and machine intelligence 40, no. 4 (2018): 1002-1014.

[4] Dhamija, Jalendu, Tanupriya Choudhury, Praveen Kumar, and Yogesh Singh Rathore. "An Advancement towards Efficient Face Recognition Using Live Video Feed:" For the Future"." In Computational Intelligence and Networks (CINE), 2017 3rd International Conference on, pp. 53-56. IEEE, 2017.

[5] Mokhayeri, Fania, Eric Granger, and Guillaume-Alexandre Bilodeau. "Domain-specific face synthesis for video face recognition from a single sample per person." IEEE Transactions on Information Forensics and Security 14, no. 3 (2019): 757-772.

[6] Zhao, Qianqian, and Hualong Cai. "The research and implementation of face detection and recognition based on video sequences." In Future Computer and Communication (ICFCC), 2010 2nd International Conference on, vol. 1, pp. V1-318. IEEE, 2010.

[7] An, Le, Bir Bhanu, and Songfan Yang. "Face recognition in multi-camera surveillance videos." In Pattern Recognition (ICPR), 2012 21st International Conference on, pp. 2885-2888. IEEE, 2012.

[8] Lu, Ke, Zhengming Ding, Jidong Zhao, and Yue Wu. "Video-based face recognition." In Image and Signal Processing (CISP), 2010 3rd International Congress on, vol. 1, pp. 232-235. IEEE, 2010.

[9] Hatimi, Hicham, Mohamed Fakir, and Mohamed Chabi. "Face recognition using a fuzzy approach and a multi-agent system from video sequences." In Computer Graphics, Imaging and Visualization (CGiV), 2016 13th International Conference on, pp. 442-447. IEEE, 2016.

[10] Hu, Changbo, Josh Harguess, and J. K. Aggarwal. "Patch-based face recognition from video." In Image Processing (ICIP), 2009 16th IEEE International Conference on, pp. 3321-3324. IEEE, 2009.

[11] Wang, Ya, Tianlong Bao, Chunhui Ding, and Ming Zhu. "Face recognition in real-world surveillance videos with deep learning method." In Image, Vision and Computing (ICIVC), 2017 2nd International Conference on, pp. 239-243. IEEE, 2017.

[12] Gong, Dihong, Kai Zhu, ZhiFeng Li, and Yu Qiao. "A semantic model for video based face recognition." In Information and Automation (ICIA), 2013 IEEE International Conference on, pp. 1369-1374. IEEE, 2013.

[13] He, Lingxiao, Haiqing Li, Qi Zhang, and Zhenan Sun. "Dynamic Feature Matching for Partial Face Recognition." IEEE Transactions on Image Processing 28, no. 2 (2019): 791-802.

[14] Su, Ya. "Robust video face recognition under pose variation." Neural Processing Letters 47, no. 1 (2018): 277-291.

[15] Yang, Shuo, Ping Luo, Chen-Change Loy, and Xiaoou Tang. "Wider face: A face detection benchmark." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5525-5533. 2016.

[16] Pandit, Trupti M., P. M. Jadhav, and A. C. Phadke. "Suspicious object detection in surveillance videos for security applications." In Inventive Computation Technologies (ICICT), International Conference on, vol. 1, pp. 1-5. IEEE, 2016.

[17] Cheng, De, Yihong Gong, Sanping Zhou, Jinjun Wang, and Nanning Zheng. "Person re-identification by multi-channel parts-based cnn with improved triplet loss function." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1335-1344. 2016.

[18] Parkhi, Omkar M., Andrea Vedaldi, and Andrew Zisserman. "Deep face recognition." In BMVC, vol. 1, no. 3, p. 6. 2015.

[19] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 815-823. 2015.

[20] Yang, Ming-Hsuan, David J. Kriegman, and Narendra Ahuja. "Detecting faces in images: A survey." IEEE Transactions on pattern analysis and machine intelligence 24, no. 1 (2002): 34-58.

[21] Yang, Ming-Hsuan. "Recent advances in face detection." IEEE ICPR 2004 Tutorial (2004).

[22] Ohlyan, Sonia, Sunita Sangwan, and T. Ahuja. "A survey on various problems & chal-lenges in face recognition." *Int. J. Eng. Technol.(IJERT)* 2, no. 6 (2013): 2278-0181.

[23] Taigman, Yaniv, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. "Deepface: Closing the gap to human-level performance in face verification." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1701-1708. 2014.

[24] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell. Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. In Proc. CVPR Workshops, 2011.