



Video Based Face Recognition: A Survey

Apurwa Khandare¹, Prof. A.S. Narote²

Student, Department of Information Technology¹, Assistant Professor, Department of Information Technology²
University of Pune, Pune, India

Abstract: Face recognition is a widely studied topic of all the times. But there are lot many areas that are still to be improved. Video based face recognition is interesting but tough to achieve. There are many methods proposed in past to achieve good and accurate results in VFR's but at the cost of time, money and lots of human efforts. Over here an effort is made to study all possible methods in the various fields of face recognition especially in Video based face recognition to know there advantages and disadvantages of each of them to know which one is better amongst them having greater accuracy and less complexity, time span and cost.

Keywords: Video face recognition, Face Recognition.

I. Introduction and Related Works

Face recognition has been a most studied topic in areas of biometric research and pattern recognition in past decades. Many algorithms are proposed in the past to achieve accurate recognition results till the date and some of them have succeeded in achieving so but only in case of still images. Whenever these algorithms are presented with the videos or video sequences either there accuracy level drops or some of them manage achieving good recognition rates but at the cost of time and extra expense which is not at all desirable[3]. Face recognition has gain interest not only because it is challenging task but also due because wide variety of applications where human identification is needed.

All face recognition algorithm consists of three major phases: Face Detection where whenever the face image appears for very first time its features are detected using algorithms like ad-boost detector and stored in the database, Face Tracking where whenever the same face occurs the track is maintained using trackers, lastly Face Recognition where test faces are identified by matching them with trained face images in database [4].

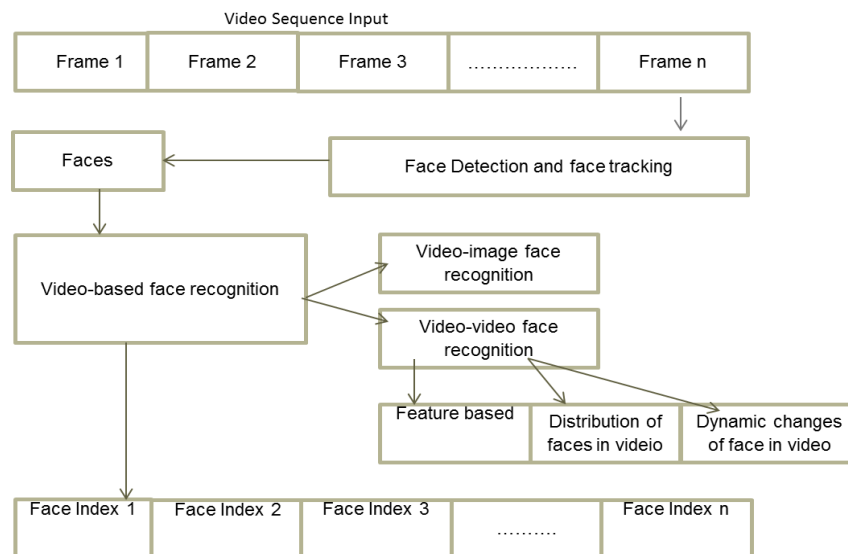


Fig.1 Modules and Classification of Video Based Face Recognition [5]

Face Recognition techniques can be classified into three types [2] and [5]:

1. Still Image based Face Recognition
2. Video-Image based Face Recognition
3. Video-Video based Face Recognition

1. Still-Image based face Recognition (SIFR):

In this type face recognition methods input as well as database both contains the still face images in other words the database contains the detected face images and when an test image is given if only the test image matches with trained still image recognition occurs. Many methods have been studied in this form of face recognition one such method is given in [6] called eigenfaces which uses principal component analysis (PCA) to efficiently represent the picture. It states that test faces can be reconstructed by weights assigned to each trained face. The weight of each face is obtained by projecting the test face image onto eigen-picture. It achieves up to 90% of

accuracy but its disadvantage is that it does not provide invariance to changes in scale and lighting conditions. The method in [7] is mainly based on the key frame selection. Over here still images are collected from internet and key frames are selected using clustering. The cluster centres are compared to test image using nearest-neighbour algorithm followed by probabilistic voting to make a final prediction. The accuracy of this method ranges as 85%-87%. This method reduces the search space but it is very complicated and cluster matching algorithm used here in many cases. One more method which uses still image based face recognition is the one in [8] which combines two standard algorithms PCA and support vector machines (SVM). It is two ways recognition based on feature vectors where PCA is used for feature extraction from an image and SVM is used for classification of those features. It is a simple method and achieves accuracy rate up to 99% which the highest among all still image based face recognition methods but it takes a little more time in comparison to other methods.

II. Video-Image Based Face Recognition (VIFR):

Over Here input to a system is a video but face images are stored in the database. That is whenever a video is given as a test video to the system the faces are extracted from it are then compared and matched with the faces stored in database at the time of training phase. It is an extension to still image based face recognition.

Method in [9] uses this type of recognition and called as Sparse Representation Based Classification (SRC). It states the principal that the test image can be represented as linear combination of face detected in a large face database. This method gives an accuracy of about 86%. Though it is a fast and efficient method the ℓ^1 -minimization used in it makes it very expensive. Another method which uses this type of recognition is called Multi-task Joint sparse representation [1] used in Movie2Comics system for face recognition aims to reconstruct a test sample with multiple features using least number of training samples possible. The method can be explained as follows:

Let given a training set with J classes in which each sample has K different modalities features such as colour, texture and shape. For each modality (task) index $k = 1, \dots, K$, denote

$$X_k = [X_{k1}, \dots, X_{kJ}]$$

The training feature matrix $X_{kj} \in \mathbb{R}^{m_k \times p_j}$ is associated with the j th class. Where m_k is dimensionality of kth modality, p_j is number of training samples in class j, and $\sum_{j=1}^J p_j = p$ is total number of training samples.

A supervised K-task linear representation problem when given a test sample with K modalities of representations is considered as

$$y_k = \sum_{j=1}^J X_{kj} w_{kj} + e_k, k = 1, \dots, K.$$

Where $w_{kj} \in \mathbb{R}^{p_j}$ is reconstruction coefficient vector associated with jth class, and $e_k \in \mathbb{R}^{m_k}$ is the residual term. $w_k = [(w_{k1})^T, \dots, (w_{kJ})^T]^T$ are representation coefficients in task k and $w_j = [w_{1j}, \dots, w_{Kj}]$ are representation coefficients from jth class across different tasks. The multi-task joint sparse representation is formulated as solution to the following problem of multi-task least square regression with $\ell_1, 2$ mixed norm regularization:

$$\min_W \frac{1}{2} \sum_{k=1}^K \|y_k - \sum_{j=1}^J X_{kj} w_{kj}\|^2 + \lambda \sum_{j=1}^J \|w_j\|_2.$$

The model used can be effectively optimized using accelerated proximal gradient method. Though it gives correct results experiments have shown that it fails to recognize users in case of multiple users in single frame resulting in accuracy of 85%.

III. Video- Video Based Face Recognition (VVFR):

In this type of recognition input as well as database is a video. Whenever a test video is given to such type of system it compares it with video sequences in database and recognizes the faces. This is the fastest and most recent techniques amongst all and it is tough to achieve.

Its further divide in to three subtypes:

- a. Feature Vector Based VVFR
- b. Distribution of Faces in Video Based VVFR
- c. Dynamic Changes of Faces in Video based VVFR.

a. Feature Vector Based VVFR:

The principal of this method is to extract features from input video which are then used to match with all the database videos. Ajmal Main et al. [10] proposed an approach based on this concept where the algorithm is mainly uses unsupervised learning. It is a two phase approach where the first phase is a learning phase in which input video clips are automatically clustered and during the second phase based on Difference-of-Gaussian function distance between clusters is found and they are classified. The other methods as in [11] & [12] proposed by Lapedriza et al. and Wolf and Shausha respectively uses this type of recognition. The one in [11] uses PCA as feature extractor and a classifier to classify the features. Whereas the one in [12] uses kernel Hiber Space for feature extraction and svm as a classifier. Feature Vector Based are known to be simplest amongst all

the methods and achieves accuracy up to 90% But faces drawback such as difficulty in making threshold setting decision.

b. Distribution of Faces in Video based VVFR:

In this faces in videos are treated as variables with some weights or probability. Similarity in the faces is computed by similarity in corresponding probability density function. The method in [13] is an example of this approach which involves a novel idea of making use of method based on appearance manifolds expressed in image space. Complex non-linear manifold is expressed as a collection of subset and by the connectivity among them. Affine plane is used to approximate each pose manifold. K-mean algorithm helps to sample and cluster constructed exemplar representation from videos. Another method in [14] where Arandjelon'c et al. made use of a kernel based method to map low dimensional space in to a high dimensional space. It uses PCA to solve complex non-linear problem in high dimensional space. The advantage of this method is it generates good results in spite of occlusion and variations. It achieves up to 86-88% accuracy. Disadvantage of this method is that it is a much generalised method and it is sensitive to illumination changes.

c. Dynamic Changes of Faces in Video based VVFR:

Here the temporal information helps to identify the facial dynamics changes and its application as identifier for person recognition in video sequences.

Liu et al. [15] makes use of an adaptive HMM (Hidden Markova Model) to perform video-based face recognition. Over here during training process, the statics of training, view sequence of each subject and the temporal dynamics are learned by an HMM. During the recognition process the temporal characteristics of the test video sequence are analysed over time by the HMM corresponding to each subject. In this method first all face images are reduced to low dimensional feature vectors by PCA. Given a face database with L subjects and each subject has a training video sequence containing T images $f_1 = \{f_{1,1}, f_{1,2}, \dots, f_{1,T}\}$ $1 \leq l \leq L$.

By performing Eigen analysis for $L \times T$ samples, an eigenspace is obtained with a mean vector m and a few eigenvectors $\{v_1, v_2, \dots, v_d\}$ which are then projected into eigenspace and corresponding feature vectors are generated. These vectors are used as training vectors for HMM which for each subject is given as: $\lambda = (A, B, \Pi)$

Where A is state transition probability matrix, B is observation probability density function and Π is initial state distribution. Adaptive HMM is nothing but during recognition once the test sequence or a subject is recognized the same can be used to update the HMM of that subject which will learn the new appearance in this sequence and provide enhance model of that subject. The decision whether to update current subject is made based on likelihood difference compared with a set threshold. Lee et al. [14] also proposed recognition by detecting dynamic changes in faces. The method is basically based on probabilistic appearance manifold approach for face tracking and recognition using video sequence. Bayesian inference was used to include temporal coherence of human motion in distance calculations. Unsupervised learning results in better modelling over time and its performance is better than majority voting used in the image-based face recognition are main advantages of this method but it faces certain drawbacks as threshold decisions, complication in computation and also it does not employ method of face tracking which can further improve recognition. It achieves recognition rate in between 85-87%.

IV. Conclusion

It has been seen through literature survey that Video-Video based face recognition is far better than the still image and Video-image based face recognition. Amongst the three types of VVFR's Feature Vector based VVFR is the most accurate with least disadvantages. PCA+ SVM feature vector based technique achieves up to 99% of accuracy in case of Still-Image based face recognition but is never used for Video face recognition. Therefore, if PCA+SVM feature vector based is used for face recognition are used for VVFR's can achieve accuracy more than 95% in case VVFR's which will be higher than any other existing methods.

References

- [1] M. Wang, R. Hong, Yuan, S. Yan, & T. Chua, 2012, "Movie2comics: Toward a Lively video content representation," IEEE transaction on multimedia, Vol.14 (3), 858-870.
- [2] A. Khandare and A. Narote, 2014, "Movie2Comics: An Effective script face mapping through PCA based face recognition," Cyber Times International Journal of Tech. and Management, Vol.7 (1), 571-576
- [3] X. Liu and T. Cheng, 2003, "Video-based face recognition using adaptive hidden Markova Models", In Computer Vision and Pattern Recognition, in Proc.IEEE Computer Society Conference on, vol.1, 1-340
- [4] S.Patil and D.Deore, 2012, "Video- based Face Recognition: A Survey", World Journal of Science and Technology, vol. 2(4), 136-139
- [5] Z.Zhang, W. Wang, and Y. Yang, 2007, "Video based face recognition: state of the art", Springer, 1-9
- [6] M. Turk and A. Pentland, 1991, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, vol.3(1), 71-86
- [7] M. Zhao, J. Yagnik, H. Adam and D. Bau, 2008, "Large scale learning and recognition of faces in web videos," FG
- [8] S.Lin, S.Kung, & L.Jin, 2008, "Support Vector Machine, PCA and LDA in face Recognition", Journal of Electronic Engg., vol.59 (4), pp. 203-209
- [9] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, 2009, "Robust face recognition via sparse representation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31 (2), 210-227
- [10] A. Mian, 2008, "Unsupervised learning from local features for video-based face recognition", 8th IEEE International Conference on Automatic Face & Gesture Recognition, 1-6

- [11] A. Lapedriza, D. Masip, and J. Vitria, 2006, "*Face verification using external feature*", In Automatic Face and Gesture Recognition, 2006, 7th International Conference, IEEE, 132–137
- [12] L. Wolf and A. Shashua. 2003, "*Kernel principal angles for classification machines with applications to image sequence interpretation*", in Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, 635-640
- [13] K. C. Lee, J. Ho, M. Yang, and D. Kriegman, 2003, "*Video-based face recognition using probabilistic appearance manifolds*," in Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 313–320
- [14] W. Fan, Y. Wang, and T. Tan, 2005, "*Video-based face recognition using Bayesian inference Model*," In Audio-and Video-Based Biometric Person Authentication, Springer, 122–130.
- [15] X. Liu and T. Cheng, 2003, "*Video-based face recognition using adaptive hidden Markov Models*", In Computer Vision and Pattern Recognition, in Proc. IEEE Computer Society Conference on, vol.1, 1–340