A survey on Video-based Face Recognition Approaches

J.SUNEETHA

1Research Scholar, S.P.M.V.V, Tirupathi, INDIA.

Abstract

Face recognition is one of the most suitable applications of image analysis. It’s a true challenge to build an automated system which equals human ability to recognize faces. While traditional face recognition is typically based on still images, face recognition from video sequences has become popular recently due to more abundant information than still images. This paper presents an overview of face recognition scenarios and video-based face recognition system architecture and various approaches are used in video-based face recognition system. This paper provides an up-to-date survey of video-based face recognition research.

Keywords: Face Recognition, Video-based, Survey.

1. INTRODUCTION

Face recognition is a biometric approach that employs automated method to verify or recognize the identity of a living person based on his/her physiological characteristics. It also used in wide range of commercial and law enforcement and interesting area in real time applications. Face recognition has several advantages over other biometric technologies: It is natural, nonintrusive, and easy to use [1]. Face recognition system can help in many ways: for example some applications are Checking for criminal records and Detection of a criminal at public place, Finding lost children’s by using the images received from the cameras fitted at some public places and detection of thief’s at ATM machines, Knowing in advance if some unknown person is entering at the border checkpoints and so on. A face recognition system can operate in either or both of two modes: (1) face verification (or authentication), and (2) face identification (or recognition). Face verification involves a one-to-one match that compares a query face image against a template face image. Face identification involves one-to-many matches that compare a query face image against all the template images in the database to determine the identity of the query face. The first automatic face recognition system was developed by Kanade[2], so the performance of face recognition systems has improved significantly.

Face recognition in videos is an active topic in the field of image processing, computer vision and biometrics over many years. Compared with still face recognition videos contain more abundant information than a single image so video contain spatio-temporal information. To improve the accuracy of face recognition in videos to get more robust and stable recognition can be achieved by fusing information of multi frames and temporal information and multi poses of faces in videos make it possible to explore shape information of face and combined into the framework of face recognition. The video-based recognition has more advantages over the image-based recognition. First, the temporal information of faces can be utilized to facilitate the recognition task. Secondly, more effective representations, such as a 3D face model or super-resolution images, can be obtained from the video sequence and used to improve recognition results. Finally, video-based recognition allows learning or updating the subject model over time to improve recognition results for future frames. So video based face recognition is also a very challenging problem, which suffers from following nuisance factors such as low quality facial images, scale variations, illumination changes, pose variations, Motion blur, and occlusions and so on[3].

Face recognition can generally be categorized into one of the following three scenarios based on the characteristics of the image(s) to be matched. Such as Still-to-still recognition, Video-to-image face recognition, Video-to-video face recognition [4]. i) Research on still image face recognition has been done for nearly half a century. Still-to-still image matching is the most common process and is used in both constrained and unconstrained applications, but it suffers from several factors those are the need to constrain the face recognition problem, computational constraints, and the large amount of legacy still face images (e.g. id cards, mug shots). ii) Video-to-image face recognition can be seen as an extension of still image based face recognition. Video-to-still image matching occurs when a sequence of video frames is matched against a database of still images (e.g. mug shots or Identification photos). The input of the system is videos while the database is still face images. Compared to traditional still image based face recognition, how to explore the multi-frame information of the input video is the key to enhance the performance. In summary, image-video based methods make use of multi-frame information to improve the accuracy of face recognition, and improve the robustness to deal with pose variations, occlusions and illumination changes.iii) Video-to-video matching, or re-identification, is performed to find all occurrences of a subject within a collection of video data. Re-identification is generally a necessary
pre-processing step before video-to-still image matching can be performed. Compared to video-image based methods, both the system input and the database this category are in the form of videos, which is a more difficult problem to solve. Based on the state of the arts, there are mainly three types of solutions this problem, those are i) Based on feature vector extracted from video input ii) Based on probability density function or manifold to depicts the distribution of faces in videos iii) Based on generative models to describe dynamic variance of face in images.

2. VIDEO-BASED FACE RECOGNITION

Video based face recognition in image sequences has gained increased interest based primarily on the idea expressed by psycho physical studies that motion helps humans recognize faces, especially when spatial image quality is low. The traditional recognition algorithms are all based on static images but video-based face recognition has been an active research topic for decades. It is categorized into two approaches those are i) Set-based and ii) Sequential-based approaches[5]. Set-based approaches consider videos as unordered collections of images and take advantage of the multitude of observations where as sequence-based approaches explicitly use temporal information to increase efficiency or enable recognition in poor viewing conditions.

2.1 SYSTEM ARCHITECTURE

Video-based face recognition systems consist of three modules: i) Face detection module ii) Feature extraction module iii) Face recognition module.

2.1.1 Face detection

Face detection is the first stage of a face recognition system. This module system takes a frame of a video sequence and performs some image processing techniques on it in order to find locates candidate face region. System can operate on static images, where this procedure is called face localization and dealing with videos procedure is called face tracking. The purpose of face localizing and extracting the face region from the background. Face detection can be performed based on several things those are skin texture, motion (for faces in videos), facial/head shape, facial appearance, or a combination of these parameters. An input image is scanned at all possible locations and scales by a sub window. Face detection is posed as classifying the pattern in the sub window as either face or non-face. Face detection techniques similar to those applied for still images are then employed to find the exact location of faces in the current frame, thus initiating face and facial feature tracking. Face tracking techniques include head tracking, where the head is viewed as a rigid object performing translations and rotations. While the tracking module finds the exact position of facial features in the current frame based on an estimate of face or feature locations in the previous frame(s). Face detection based on videos consists of generally three main processes. Firstly frame based detection, in this process, lots of traditional methods[6] for still images can be introduced such as statistical modeling method, Neural network-based method, SVM-based method, HMM method, BOOST method and Color-based face detection, etc. Secondly combination of detection and tracking, this says that detecting face in the first frame and then tracking it through the whole sequence. Since detection and tracking are independent and information from one source is just in use at one time, loss of information is unavoidable. Finally, instead of detecting each frame, temporal approach exploits temporal relationships between the frames to detect multiple human faces in a video sequence.
2.1.2 Feature extraction

The extraction of discriminant features is the most fundamental and important problem in face. After obtaining the image of a face, the next step is to extract facial features. There are two types of features can be extracted i) Geometric features ii) Appearance Features. Geometric features represent the shapes and location of facial components such as eyebrows, eyes, nose, mouth etc. Experimental results exhibited that the facial features cannot always be obtained reliably because the quality of images, illumination and other disturbing factors. The Appearance based features present the appearance (skin texture) changes of the face, such as Wrinkles and furrows.

2.1.3 Face recognition

Face recognition is the most significant stage in the entire system. Videos are capable of providing more abundant information than still image. The major advantages for using videos are firstly the possibility of employing redundancy contained in the video sequence to improve still images recognition performance, second dynamic information is available and thirdly to improve recognition effects from the video sequence using more effective representations such as a 3D face model or super-resolution images. Finally video-based recognition allows learning or updating the subject model over time. Though the advantages are obvious, there also exits some disadvantages. For example, poor video quality, low image resolution, and other influence factors (such as illumination, pose change, motion, occlusion, decoration, expression, large distance from camera, etc). The face recognition methods divided into two categories such as i) Frame-based recognition ii) Sequence-based recognition. The Frame-based recognition method is based on static images and sequence-based recognition method is based on dynamic video images. Sequence-based Expression recognition uses the temporal information of the sequence to recognize the expressions for one or more frames. Hidden Markov models (HMM), recurrent neural networks and rule based classifiers use sequence-based Expression Recognition. Sequence-based Expression Recognition classification schemes divided into two types such as dynamic and static classification. The static classifiers are classifiers that classify a frame in the video to one of the facial expression categories based on the tracking results of that frame. Mainly based on Bayesian network and Gaussian Tree-Augmented Naïve (TAN) . Bayes classifiers. Dynamic classifiers are classifiers that take into account the temporal pattern in displaying facial expression. A multi-level HMM classifier is used for dynamic classification. The fig 2 represents the example of Video-based face recognition.

![Input Video](image)

Fig. 2. Example of a Video-based face recognition.

3. APPROACHES OF VIDEO-BASED RECOGNITION

There are various aspects of approaches for video based face recognition such as i) Spatio-temporal information based approaches ii) Statistic model based approaches iii) Hybrid cues based approaches iv) Advanced Topics

3.1 Spatio-temporal information based approaches

Most of the recent approaches utilize spatio-temporal information for face recognition in video. So some [7] use temporal voting to improve identification rates. There are also using several algorithms which extract 2D or 3D face structure from the video. The distance between two videos is the minimum distance between two frames across two videos. Zhou and Chellappa presented a sequential importance sampling (SIS) method to incorporate temporal information in a video sequence for face recognition [8], it nevertheless considered only identity consistency in temporal domain and thus it may not work well when the target is partially occluded. In [9], Krueger and Zhou selected representative face images as
exemplars from training videos by on-line version of radial basis functions. This model is effective in capturing small 2D motion but may not deal well with large 3D pose variation or occlusion. The condensation algorithm could be used as an alternative to model the temporal structures [10].

The methods based on spatio-temporal representations for face recognition in video have some drawbacks: i) The local information is very important to facial image analysis, it is not well exploited ii) personal specific facial dynamics are useful for discriminating between different persons, however the intra-personal temporal information which is related to facial expression and emotions is also encoded and used and iii) equal weights are given to the spatiotemporal features despite the fact that some of the features contribute to recognition more than others iv) a lot of methods can only handle well aligned faces thus limiting their use in practical scene[11].

3.2 Statistic model based approaches
Zhou et al. [12] obtained statistical models from video by using low level features (e.g., by PCA) contained in sample images, which was used to perform matching between a single frame and the video stream or between two video streams. The mutual subspace method in [13] took the video frames for each person separately to compute many individual eigen spaces, considering the angle between input and reference subspaces formed by the principal components of the image sequences as the measure of similarity. Principal component null space analysis (PCNSA) is proposed in [14], which is helpful for nonwhite noise covariance matrices. Recently, the Autoregressive and Moving Average (ARMA) model method is proposed in [15] to model a moving face as a linear dynamical object. S. Soatto, G. Doretto, and Y. Wu proposed dynamic textures for video-based face recognition. Kim et al. applied HMM to solve the visual constraints problem for face tracking and recognition [16].

3.3 Hybrid cues
Video can provide more efficient information than still image. Some methods utilize other cues obtained from video sequences, such as voice, gait, motion etc. In[17] Shan et al. investigated the fusion of face and gait at feature level and gained performance increase by combining the two cues[18] adopted a face and speaker recognition techniques for audio-video biometric recognition. The paper combined histogram normalization, boosting technique and a linear discrimination analysis to solve the problem like illumination, pose and occlusion and proposes an optimization of a speech denoising algorithm on the basis of Extended Kalman Filter(EKF). In [19], an approach was presented by radial basis function neural networks, which is used to recognize a person in video sequences by using face and mouth modalities.

3.4 Advanced Topics
The more popular areas of video-based face recognition technology for the past several years are as follows. Such as illumination, pose, 3D researches, Low Resolution.

i) Illumination
It is difficult for system to make recognition of individuals when change in light is larger. Belhumeur et al. and Bartlett et al. adopted the PCA by discarding the first few principal components and achieved better performance for images under different lighting conditions. Recently, in [20], an effective method of handling illumination variations was presented by using illumination cone. This method also dealt with shadowing and multiple lighting conditions. which was on the basis of 3D linear subspace. The main side effect of this method is that the training set requires more than 3 aligned images per person.

ii) pose
Pose is another most important factor for face recognition system. Current approaches can be divided into three groups: multiple images approaches, hybrid approaches and single image based approaches. New AAM methods have been proposed to handle both varying pose and expression. Using Eigen-light-fields and Fisher light-fields method was proposed to do pose invariant face recognition. A method by 3D model of the entire head for exploiting features like hairline, which handled large pose variations in head tracking and video-based face recognition was presented [21].

iii) 3D researches
Face recognition based on 3D is a hot research topic. Generally, reprehensive methods can be divided into three main categories, namely, 2D images based, 3D images based and multimodal systems. The differences among these three categories are as follows: the first category includes approaches which use 2D images and 3D generic face model to improve the robustness and recognition rate. And for the second one, the methods work directly on 3D datasets. While the last group means those which utilize both 2D and 3D information. Since 2000, more and more multimodal approaches have been proposed to improve face recognition performance. Dalong Jiang et al. [22] proposed an efficient and fully automatic 2D-to-3D integrated face reconstruction method in an analysis-by-synthesis manner.

iv) Low Resolution
It is difficult to recognize human faces in video of low resolution. With the widely use of camera (surveillance etc.), solutions which solve such problems achieve more and more attention. The main two methods are Super Resolution (SR) and Multiple Resolution-faces (MRF) approach. In [23], color invariance was applied to face recognition. Their result showed that color invariants do have substantial discriminative power and increase the robustness and accuracy for low resolution facial recognition.

4. VIDEO-VIDEO FACE RECOGNITION

The Image-video based methods make use of multi-frame information to improve the accuracy of face recognition, and improve the robustness to deal with pose variations, occlusions and illumination changes. In this method both the system input and the database are in the form of videos, which is a more difficult problem to solve. So using video-video face recognition there are mainly three types of solutions of this problem, which are listed as follows:

1. Based on feature vector extracted from video input
2. Based on probability density function or manifold to depict the distribution of faces in videos
3. Based on generative models to describe dynamic variance of face in images

4.1 Feature vector based methods

The basic idea of this solution is to extract feature vectors from input videos, which are used to match with all the videos in the database. To extract the feature vector the approaches are namely, feature based (e.g. Bottom-Up) and image based (e.g. Appearance-Based and Template matching) approaches. Features based approaches extract facial features from an image and manipulate its parameters such as angles, size, Occlusions and distances. Image base approaches rely on training and learning set of examples of objects of interest. However, dealing with video introduces other approaches for face detection such as motion based approach.

4.1.1 Knowledge-based (Top-Down) approach

In this approach, the relationship between facial features is captured to represent the contents of a face and encode it as a set of rules. Coarse-to fine scale is used in lots of algorithms classified under this category, in which the coarsest scale is searched first and then proceeds with the others until the finest scale is reached.

4.1.2 Feature invariant (Bottom-Up) approach

In this approach, the face's structural features which do not change under different conditions such as varying view points, Scale, pose angles and/or lightning conditions. Common algorithms used under this category are: Color-based approach, or so called skin-model based approach. This approach makes use of the fact that the skin color can be used as indication to the existence of human using the fact that different skins from different races are clustered in a single region.

4.1.3 Facial features based approach

This approach, in which global (e.g. skin, size, and shape) and/or detailed (e.g. eyes, nose, and lips) features are used, has become popular recently. Mostly, the global features first are used to detect the candidate area and then tested using the detailed features. Texture The human face differs from other objects in texture. This method, examines the likelihood of sub image to belong to human face texture, using Space Gray Level dependency (SGLD) matrix.

In [24], Horst Eidenberger proposed a Kalman filter-based method to describe invariant face features in videos based on a compact vision model, which achieves high performance in the some databases. Park et al. [25] created multiple face templates for each class of face in database according to the video information, dynamic fuse the multiple templates as the feature. Lapedriza et al. [26] built PCA feature subspaces for the faces of the input and database, which achieves recognition based on the distance between the subspace measured by geometric angles. In order to remove the effect of light, gesture, motion blur and facial expressions, Fukuiand Yamaguchi [27] projected the feature space to the constraint subspace. This kind of solution makes use of multi frames of the input video to obtain a discriminant feature representation. However, the spatial information of input videos is neglected, which limits the performance of feature vector based approaches.

4.2 Distribution of faces in videos

The main idea of this category is to treat faces in videos as random variables of certain probability. The similarity of faces are measured by similarity of corresponding probability density distributions. In this using Template matching methods are based on measuring the degree of similarity between the candidate sub image and the predefined stored face pattern. The predefined image might be for the whole face pattern or the individual face features such as eyes, nose and lips. Common algorithms used under this category are: Predefined face templates, in which several templates for the whole, individual or both (whole and individual) parts of a face are stored. Deformable Templates in which an elastic facial feature model as a reference model where the deformable template mode of the object of interest, is fitted in.
In [28], each face in the database and input video are modeled with GMM, and Kullback-Leibler divergence is measured as the similarity measurement to achieve recognition. In [29], Arandjelović et al. made use of kernel-based methods to map low-dimensional space to high-dimensional space, and then use low-dimensional space of linear methods (such as PCA) to solve complex nonlinear problems in the high-dimensional space. Zhou et al. [30] mapped the vector space into RKHS (Rep reducing Kernel Hilbert Space) by kernel-based methods to calculate the distance between the probability distribution. In a multi-view dynamic face model is built to achieve face recognition. Firstly, dynamic face model are constructed including a 3D model, a texture model and an affine change model. Then a Kalman filter is adopted to obtain the shape and texture, which builds a segmented linear manifold for each single person with the face texture reduced by KDA (Kernel Discriminate Analysis). Face recognition is achieved in the following by trajectory matching. However, the 3D model estimation requires a lot of multi-angle images and a larger complexity computational. This kind of solution is much better than the feature vector based solution, which makes use of probability theory to enhance the performance. However, the dynamic change information of faces in videos is neglected, which has potential to improve the video based face recognition.

4.3 Dynamic changes of faces in videos

The temporal information in video sequences enables the analysis of facial dynamic changes and its application as a biometric identifier for person recognition. We have utilize the human nature that human will have at least small amount of movements such as eyes blinking and/or mouth and face boundary movements. We can get this information easily because dealing with video sequence by which the whole sequence of the object's movements can be obtained. Taking that point into account we can reduce the error that occurs due to false detection of a human face and minimize the time of simulation.

Matta et al. proposed a multi-modal recognition system [31,32]. They successfully integrated the facial motion information with mouth motion and facial appearance by taking advantage of a unified probabilistic framework. In [33], Huang and Trivedi developed a face recognition system by employing HMMs for facial dynamic information modeling in videos. Each covariance matrix was gradually adapted from a global diagonal one by using its class-dependent data in training algorithms. Afterwards, Liu and Cheng [34] successfully applied HMMs for temporal video recognition (as illustrated in Fig. 4) by improving the basic implementation of Huang and Trivedi. Each test sequence was used to update the model parameters of the client in question by applying a maximum a posteriori (MAP) adaptation technique.
5. CONCLUSION
Recent rapid progress of communication technology and computer science has made video-based face recognition acts as a vital role in human-machine interface and advanced communication. The main objective of this paper describes a survey of video-based face recognition modules & approaches. Still-to-Still, Video-to-Still based methods only exploit less and physiological information of the face but in Video-to-Video based methods have more and abundant information. In future video-based face recognition has made great challenge and to adopted in real application.

REFERENCES


J.SUNETHA received the B.Tech. Degree in Computer Science from Jawaharlal Nehru Technological University, Hyderabad and M.Tech. Degree in Computer Science from Sathyabhama University, Chennai. She is Research Scholar at Sri padmavathi Mahila Viswa Vidyalayam,Tirupathi, Andhra Pradesh, India.